نموذج ترخيص

أنا الطالب : معين سلمان سليم النصراوين أمنح الجامعة الأردنية و/ أو من تفوضه ترخيصاً غير حصري دون مقابل بنشر و / أو استعمال و / أو استغلال و / أو ترجمة و / أو تصوير و / أو إعادة إنتاج بأي طريقة كانت سواء ورقية و / أو إلكترونية أو غير ذلك رسالة الماجستير / الدكتوراه المقدمة من قبلي وعنوانها.

مقارنة أربعة نماذج لمعالجة التخمين في فقرات الاختيار من متعدد في إطار النموذج اللوجستي ثلاثي المعلمة وأثرها على دقة تقدير معلمة القدرة ومعالم الفقرات.

وذلك لغايات البحث العلمي و / أو التبادل مع المؤسسات التعليمية والجامعات و / أو لأي غاية أخرى تراها الجامعة الأردنية مناسبة، وأمنح الجامعة الحق بالترخيص للغير بجميع أو بعض ما رخصته لها.

اسم الطالب: معين سلمان سليم النصراوين

التوقيع: [توقيع]

التاريخ: ١٢/٥/٢٠١٥م

مقارنة أربعة نماذج لمعالجة التخمين في فقرات الاختيار من متعدد في إطار النموذج اللوجستي ثلاثي المعلمة وأثرها على دقة تقدير معلمة القدرة ومعالم الفقرات

إعداد

معين سلمان سليم النصراوين

المشرف

الأستاذ الدكتور "محمد وليد" موسى البطش

قدمت هذه الأطروحة استكمالاً لمتطلبات الحصول على درجة الدكتوراه في

علم النفس التربوي/ القياس والتقويم

كلية الدراسات العليا

الجامعة الأردنية

نيسان ٢٠١٥ م

قرار لجنة المناقشة

نوقشت هذه الأطروحة (مقارنة أربعة نماذج لمعالجة التخمين في فقرات الاختيار من متعدد في إطار النموذج اللوجستي ثلاثي المعلمة وأثرها على دقة تقدير معلمة القدرة ومعالم الفقرات) وأجيزت بتاريخ 2015/4/21.

أعضاء لجنة المناقشة	التوقيع
الأستاذ الدكتور "محمد وليد" موسى البطش، مشرفاً	....
أستاذ / قياس وتقويم وإحصاء	
الدكتور حيدر إبراهيم ظاظا، عضواً	....
أستاذ مشارك/ قياس وتقويم	
الدكتورة فريال محمد أبو عواد، عضواً	....
أستاذ مشارك/ قياس وتقويم	
الأستاذ الدكتور يوسف محمد سوالمة، عضواً خارجياً	....
أستاذ / قياس وتقويم (جامعة اليرموك)	

الإهداء

ثمة حياة تأتيك على هيئة بشر ...

لو أردت أن تكتب عنهم أو لهم، تستحضر كل ما لديك من وهج ليضيء

بحروف أسمائهم ....

" أبي " لروحك في عليائها ... حيث الرائعون يعيشون قليلاً جدًا لكن عطرهم يظل إلى ما وراء

الأبد ...

" أمي " زوادة الروح و زيتونة البيت الوارفة وعطر الجنة، جميلة الجميلات عند عتبات قلبها

مبتدأ العمر ومنتهى الحياة ...

" أشقائي " ثلاثة أقمار تنير درب العمر ...

" شقيقاتي " خمس شموس تسطع بدفء لا ينتهي...

" زوجتي " بستان الورد في قلبي، من عينيها تنهمر حكايات الحب

ولروحها تحج الأمنيات ...

أبناء قلبي " عون " و " لين " وتذوب الحروف خجلًا لإبتسامة قلبيهما ...

عمي وعمتي صاحبا القلب الكبير ...

لهم جميعًا

أهدي هذا المجهود المتواضع...

معين سلمان سليم النصراوين

شكر وتقدير

بعد شكر المولى عز وجل الذي أعانني على إتمام هذا العمل، أتقدم بخالص الشكر للأستاذ الدكتور محمد وليد موسى البطش المشرف على هذه الأطروحة.

كما اتقدم بالشكر لجميع أساتذتي في الجامعة الأردنية / كلية العلوم التربوية- قسم علم النفس التربوي، واخص بالذكر الدكتور حيدر ظاظا والدكتورة فريال ابو عواد على تعاونهم وتقديمهم العون والمساعدة.

كما ولا يفوتني شكر كل من أسهم بهذا العمل، واخص بالذكر الدكتورة فايزة سعادة والاستاذ طارق لافي الرواشدة على جهودهم بتدقيق ومراجعة الرسالة لغويًا، كما اتوجه بالشكر إلى الاستاذ الدكتور (باول باريت) رئيس قسم علم النفس في جامعة Aucland في نيوزلندا على جهوده بتزويدي بالدراسات والمقالات التي كنت اطلبها منه باستمرار طيلة العمل بهذه الأطروحة، كما واشكر الدكتورة والباحثة (كاو جنغ) من قسم العلوم الاحصائية في جامعة Southern Methodist في الولايات المتحدة الامريكية على مساعدتها لي في كتابة برامج التشغيل لنموذجين من نماذج هذه الأطروحة وتقديمها المشورة حينما كنت اطلبها. واخيرًا يشرفني أن أتقدم بخالص شكري وتقديري إلى أعضاء لجنة المناقشة الذين تجشموا عناء القراءة والتدقيق ممّا سيُسهم باثراء هذه الأطروحة وتحسينها، فلهم مني جزيل الشكر والعرفان.

معين سلمان سليم النصراوين

فهرس المحتويات

فهرس المحتويات

قائمة الجداول

قائمة الجداول

قائمة الجداول

قائمة الجداول

قائمة الاشكال

قائمة الملاحق

قائمة المختصرات الواردة في الاطروحة

الاختصار	المعنى
3PL	النموذج اللوجستي الثلاثي المعلمة
RMSE	الجذر التربيعي لمتوسط مربع الخطأ
ARRG	نموذج إزالة التخمين العشوائي المبني على القدرة
RTE	نموذج جهد وقت الاستجابة
SB	سلوك الحل
EMIRM	نموذج الجهد المعدل لزمن الاستجابة عن الفقرة.
MLE	طريقة الارجحية العظمى في التقدير
SEE	الخطأ المعياري للتقدير
MAP	أسلوب التقدير البعدي الأعظم
$I_i(\theta)$	دالة معلومات الفقرة
$I(\theta)$	دالة معلومات الاختبار
RE	الكفاءة النسبية
K-S	الاختبار الإحصائي كولموغروف - سمرنوف

مقارنة أربعة نماذج لمعالجة التخمين في فقرات الاختيار من متعدد في إطار النموذج اللوجستي ثلاثي المعلمة وأثرها على دقة تقدير معلمة القدرة ومعالم الفقرات.

إعداد

معين سلمان نصراوين

المشرف

الأستاذ الدكتور محمد وليد البطش

ملخـــــّص

هدفت هذه الدراسة إلى الوقوف على فاعلية أربعة نماذج لمعالجة التخمين في الاختبارات الموضوعية/الاختيار من متعدد ضمن إطار النموذج اللوجستي الثلاثي المعلمة وأثرها على تقدير معلمة القدرة ومعالم الفقرة باختلاف متغير طول الاختبار (٢٥ فقرة و٥٠ فقرة) وحجم عينة المفحوصين (٥٠٠ مفحوص و١٥٠٠ مفحوص)، والنماذج المستخدمة من نظرية الاستجابة للفقرة في هذه الدراسة هي: النموذج اللوجستي الثلاثي المعلمة، والنموذج الثنائي المعدل بإضافة قيمة ثابتة (٠,٢٠)، ونموذج عتبة التخمين، والنموذج القائم على الصعوبة.

ولتحقيق غرض الدراسة؛ تم توليد البيانات التي تمثل استجابات المفحوصين باستخدام أسلوب المحاكاة بحيث تكون مُطابقة لافتراضات ومتطلبات النموذج اللوجستي الثلاثي المعلمة باستخدام برنامج (Wingen3) وذلك وفقًا للخصائص التالية: وزعت معلمة القدرة للمفحوصين ومعلمة الصعوبة توزيعًا طبيعيًا بمتوسط (صفر) وانحراف معياري (واحد)، كذلك توزعت معلمة التمييز توزيعًا منتظمًا بقيمة دنيا تساوي(٠,٤٠) لوجت وقيمة عظمى مساوية لـ (١,٧) لوجت، أما معلمة التخمين فقد وُلدت بقيمة دنيا تساوي (٠,٠٥) وقيمة عظمى مساوية لـ (٠,٤٠)، وقد تم التحقق من افتراض أحادية البعد للبيانات المولدة التي تمثل إجابات المفحوصين باستخدام إجراءين هما معامل الاتساق الداخلي/إحصائيات الفقرة باستخدام معادلة كرونباخ الفا، وإجراء التحليل العاملي باستخدام طريقة المكونات الرئيسة، كما تم استخدام الاختبار الإحصائي (كولمجروف-سمرنوف) للتأكد من التوزيع الطبيعي لكل من معلمة القدرة المولدة ومعلمة صعوبة الفقرة المولدة. واستخدم الباحث برنامج (Winbug) والذي يعتمد طريقة التقدير البيزية لتقدير معالم الفقرات ومعلمة القدرة للمفحوصين بحسب النماذج الأربعة المستخدمة في هذه الدراسة لمعالجة التخمين، وقد مرت عملية تقدير معالم الفقرات ومعلمة القدرة للمفحوصين في ثلاث مراحل حتى تم الوصول للتقديرات النهائية.

وللإجابة عن أسئلة الدراسة المتعلقة بمدى فعالية النماذج المستخدمة في هذه الدراسة لمعالجة التخمين تم تحليل قيم معالم الفقرات ومعلمة القدرة للمفحوصين من خلال بعض مؤشرات دقة التقدير وهي: الجذر التربيعي لمتوسط مربع الخطأ (RMSE)، والتحيز (Bias)، ومؤشر الكفاءة النسبية والمتوسطات الحسابية والانحرافات المعيارية للخطأ المعياري للتقدير والتحيز، كما تم استخدم تحليل التباين للقياسات المتكررة من خلال الرزمة الإحصائية للعلوم الاجتماعية(SPSS).

أظهرت النتائج التي توصلت إليها الدراسة ان الفروق بين المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة كانت دالة عند مستوى الدلالة (α=0.05) في حالة استخدام النموذج اللوجستي الثلاثي المعلمة مع كل نموذج من النماذج الأخرى؛ الأمرالذي يُشير إلى دقة النموذج اللوجستي الثلاثي المعلمة. كما أظهرت نتائج الدراسة أن الفروق بين المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ والتحيز في حالة كل من معلمة التمييز، ومعلمة الصعوبة، ومعلمة التخمين كانت دالة إحصائيا عند مستوى الدلالة (α=0.05) في حالة النموذج اللوجستي الثلاثي المعلمة مقارنة مع النماذج الأخرى، يليه النموذج الثنائي المعدل ثم نموذج عتبة التخمين، وأخيرًا النموذج القائم على الصعوبة، أما فيما يتعلق بمؤشر الكفاءة النسبية فقد أظهرت النتائج عمومًا افضلية لكل من النموذج اللوجستي الثلاثي المعلمة، والنموذج الثنائي المعدل مقارنة مع النماذج الأخرى المستخدمة لمعالجة التخمين.

كما أظهرت نتائج الدراسة أن قيم الخطأ المعياري في التقدير والجذر التربيعي لمتوسط مربعات الخطأ، ومؤشر التحيز جميعها تقل بزيادة حجم العينة من (٥٠٠) مفحوص إلى (١٥٠٠) مفحوص، وبزيادة طول الاختبار من (٢٥) فقرة إلى (٥٠) فقرة.

الكلمات المفتاحية: التخمين، دقة التقدير، حجم العينة، طول الاختبار، النماذج المعدلة، النماذج المختلطة.

الفصل الأول

خلفية الدراسة وأهميتها

الأول

مشكلة الدراسة وأهميتها

مقدمة

تحظى عمليات القياس وأدواتها باهتمام المختصين في العلوم الطبيعية والسلوكية والإنسانية على حد سواء، فهذه العلوم جميعها تسعى إلى تطوير أساليب دقيقة وموضوعية لقياس الظواهر المتعلقة بها من أجل فهمها وتفسيرها والتنبؤ بالعلاقات القائمة بين متغيراتها، في محاولة لضبطها والتحكم بها، مما يؤدي إلى دقة في نتائجها التي تخضع لأسس ومعايير موضوعية بعيدة عن الآراء والأحكام الفردية الذاتية التي يصعب عزل تأثيرها عن الظاهرة، إذ يُقاس تقدم أي علم من العلوم بدرجة الدقة التي يصل إليها قياس الظواهر في ذلك العلم، إلا أن القياس التربوي والنفسي أكثر صعوبة من القياس الطبيعي؛ لأنّ طبيعة الظواهر التربوية والنفسية متداخلة ومتشابكة وتتأثر بالعديد من المتغيرات بشكل مباشر أو ضمني، فلا يمكن دراستها بمعزل عن تأثير هذه المتغيرات؛ الأمرالذي يتطلب جهودًا بحثية متواصلة، وأدوات قياس خاصة تختلف عن تلك الأدوات المستخدمة في العلوم الطبيعية.

وتُعّد عملية التقويم عملية قديمة قدم الإنسان فهو يُمارسها في حياته اليومية من خلال إعطاء وزن للظواهر من حوله، وإصدار أحكام تنعكس آثارها على الأفراد والموضوعات والأحداث، أما القياس فينظر إليه على أنه أداة من أدوات التقويم فهو لا يتأتّى بدون قياس؛ فكلاهما مترابطان حيث أن اجراءات عملية التقويم تشمل بصورة ضمنية إجراءات عملية القياس، ولذلك فإن مفهوم التقويم يُعّد أعم وأشمل من مفهوم القياس، ولعل التطور في التقويم وأدواته عملية ضرورية لتكون الأحكام الصادرة أكثر صدقًا وثباتًا وأقل خطأ وتاثرًا بقيم الفرد الذاتية، وفي ميدان التربية والتعليم يُعّد المتعلم المحور الأساسي في العملية التعليمية التعّلمية، ومن هنا فإن الحديث عن القياس والتقويم التربوي ينصّب بشكل أساسي وبصورة مباشرة أو غير مباشرة على نواتج التعلم لدى المتعلم، ومن يُتابع تطور عمليات القياس والتقويم التربوي يجد أن مهامه بدأت تتسع وتتشعب لتشمل جميع جوانب العملية التربوية ذات العلاقة بالتعلم؛ مثل: المنهاج، والكتاب المدرسي، وطرق التدريس التي يتبعها المعلمون، وبرامج تفريد التعليم، والبيئة التعليمية؛ ممّا يُبرز شمولية وأهمية التقويم التربوي في كافة المجالات(Mehrens and Lehmann, 1991).

لقد وجُد ميدان القياس النفسي والتربوي عمومًا من أجل توظيف مخرجات عمليات القياس والتقويم في تنظيم العمليات والبرامج التربوية ونواتجها، والإبقاء على نظام المراقبة في المؤسسات التعليمية، وتوفير قاعدة للمعلومات من أجل بناء الاستراتيجيات والسياسات الخاصة بالعملية التعليمية التعلمية، وتوفير أدوات ومؤشرات لأغراض المتابعة طويلة الأمد أو قصيرة الأمد، وذلك لضبط الجودة ومراقبة أداء القائمين على تنفيذ البرامج التربوية ومساءلتهم، بالإضافة إلى المساعدة في تلبية الحاجات التربوية الخاصة بمجتمع الطلبة في المؤسسة التربوية؛ وهناك أسباب عدة للاعتقاد بأن قيمة التحصيل التربوي أساسية ومهمة جدًا من أجل جعل عملية التربية فاعلة، إذ تُسهم عمليات القياس والتقويم في توفير المعلومات اللازمة للتعرف بين فترة وأخرى على مدى نجاح المدرسين في جهودهم حتى يقرّروا الاستمرارية أو التوقف للتطوير في ممارساتهم التعليمية أو تغييرها، فالتقويم التربوي يعمل على تزويدهم بهذه المعلومات الهامة. ومن وظائف التقويم التربوي المختلفة قياس فاعلية الخطط التدريبية، وبيان مدى كفاءتها ومقدار تحقيقها للأهداف المطلوبة، وإبراز نواحي الضعف والقوة فيها، كما يُستخدم لقياس كفاية البرامج التدريبية في البيئة التعليمية التعلمية، والتحقق من مدى نجاحها في تحقيق أهدافها المرسومة، ومن خلال التقويم تقُاس كفاية المتدربين ومدى التغيير الذي أحدثه التدريب فيهم، وكذلك تُقاس كفاية المدربين الذين قاموا بتنفيذ العمل التدريبي، والوقوف على جوانب القوة والضعف لديهم من أجل التطوير أو التحسين (Miller,Linn and Gronlund, 2009).

ويُقاس مدى نجاح عملية التقويم بمدى القيام بالوظائف والأدوات المنوطة بها، كما أنها تتطلب من القائمين عليها أن يراعوا خاصيتي الشمولية والاستمرارية لهذه العملية، وتتطلب الشمولية التنوع في أدوات القياس والتقويم لجمع المعلومات الضرورية، أما الاستمرارية فهي تشير إلى أنّ عملية التقويم لا تتوقف عند حد معين؛ فالمعلم يجمع المعلومات والملاحظات عن أداء الطالب في غرفة الصف، وربما خارج غرفة الصف بصورة مستمرة، بمعنى أن نهاية عملية تقويم معينة قد تكون بداية لعملية تقويم جديدة. ومن جهة أخرى فإن دقة عملية التقويم والفائدة منها تعتمد على دقة الأدوات التي تم توظيفها في هذه العملية، وما تمخض عنها من دقة للبيانات التي جُمعت من خلالها لاتخاذ قرار حولها؛ فالحصول على المعلومات والبيانات التي تتسم بالمصداقية والثبات يُشكّل حجر الزاوية في عملية القياس تمهيدًا لتقويمها وإصدار القرارات والأحكام السليمة (البطش، ٢٠٠٥).

ولعل المتتبع للأدب النظري والتجريبي المتعلق بالقياس والتقويم عمومًا يلاحظ أن الاختبارات التحصيلية هي إحدى الأدوات الرئيسة الهامة التي يلجأ من خلالها العاملون في المؤسسات التربوية لإثارة الدافعية لدى طلابهم وتقييم ما حققوه من نتاجات التعلم، مما يدفع هؤلاء الطلبة للمثابرة وبذل المزيد من الجهود لرفع التحصيل، كما ويُلاحظ أن هذه الاختبارات تُستخدم لأهدافٍ متنوعة: كاختيار شخص لوظيفة ما من بين مجموعة من المتقدمين لهذه الوظيفة، أو لأغراض التصنيف، كتحديد مسار الطلاب الأكاديمي بما يتناسب مع قدراتهم ومهاراتهم. كما تستخدم الاختبارات التحصيلية من قبل الجمعيات المهنية المختلفة لمنح الشهادات الخاصة بالكفاءة لممارسة مهنة معينة أو القيام بعمل ما، وتلجأ المؤسسات التربوية ضمن نظامها التربوي إلى عكس مدى النجاح الذي حققته من خلال نتائج الاختبارات التحصيلية وتعريف أفراد المجتمع بطبيعة الحاجات التربوية لهذه المؤسسات، كذلك تستخدم الاختبارات التحصيلية كوسيلة لتشخيص الصعوبات وتصميم البرامج العلاجية للأفراد، ولعل تحقّق كل ما سبق في وظائف الاختبارات التربوية واستخداماتها مرهون بدقة هذه الأدوات وثبات وصدق وموضوعية النتائج المتمخضة عنها (Poham, 2003).

وتُعّد الدقة والموضوعية والتمتع بثبات وصدق النتائج المتمخضة عن أدوات قياس التحصيل مرهونة بأنماط وأنواع هذه الأدوات، وتوجد أنماط مختلفة من الأساليب الاختبارية التحصيلية التي يمكن للعاملين في المؤسسات التربوية استخدامها: فهناك الاختبارات الشفوية واختبارات المقال، والاختبارات الموضوعية بمختلف أنواعها ولكل منها خصائصها ومميزاتها والمواقف التي يُفضل أن تُستخدم فيها؛ إذ يُنظر إلى الاختبارات المقالية على أنها ذلك النوع من الأسئلة الذي يتطلب من المفحوص تحرير إجابته وكتابتها بناءً على افتراض أن لديه حصيلة كبيرة من المعرفة، ليستدعي من خلالها المعلومات التي تتطلبها الإجابة عن السؤال، فيعمل على تنظيمها وترتيبها بشكل يُمكّنه من الإجابة، ويتم إعطاء الدرجات عند تصحيح إجابات الطالب على هذا النمط من الأسئلة بناءً على الانطباع العام الذي يخرج به المصحح عند قراءة إجابة الطالب، وبناء على توفر عناصر معينة فيها، أما الاختبارات الموضوعية فهي تُعدّ الأحدث ظهورًا مقارنة مع الاختبارات الإنشائية حيث يُشير ميريت (Merritt, 2006) إلى أن أول من وضع هذا النمط من الاختبارات هو (فريديرك كيلي) في جامعة كنساس في الولايات المتحدة الأميركية عام(١٩١٤م)؛ بهدف تقييم القدرات العقلية للجنود المشاركين في الحرب العالمية الأولى، وقد لاقت قبولاً واسعًا من قبل العاملين في المجال التربوي، وانتشر استخدامها في المؤسسات التربوية؛ وتُعدّ اليوم من أكثر أدوات قياس التحصيل استخدامًا.

وساهم انتشار أجهزة الحاسوب واستخدامها في إعداد الاختبارات وتقديمها بشكل كبير في ازدياد الإقبال على استخدام الاختبارات الموضوعية/الاختيار من متعدد وجعلها أكثر أنواع الاختبارات تداولاً في شتى المجالات (Miller,Linn and Gronlund, 2009). وقد كان لهذه الاختبارات مساهمة كبيرة في تنشيط حركة القياس التربوي والنفسي؛ وذلك بسبب كفاءتها وتعدد استعمالاتها؛ لما تتصف به من خصائص ومميزات؛ إذ تساهم في تقليل كل من وقت الإجابة من جانب الطلبة، ووقت التصحيح من جانب المدرسين، وتُعتبر أكثر سهولة وموضوعية في التصحيح مقارنة بالاختبارات المقالية، وهي كذلك قادرة على تمثيل المحتوى الدراسي للمادة التعليمية بصورة أفضل من الأنواع الأخرى من الاختبارات، وتمتاز بثبات مرتفع ويمكنها مساعدة الطلبة في التقييم والكشف عن المعارف والمهارات والقدرات التي اكتسبوها، حيث يُمكن من خلال هذا النوع من الاختبارات قياس الأهداف البسيطة والمركبة في مختلف المواضيع الدراسية، وبإختلاف المراحل التعليمية (Ebel and Frisbie, 1991;Sternberg and Grigorenko, 2002).

إنّ شيوع فكرة الاختبارات الموضوعية أدى إلى التعدد في الأشكال التي تتخذها هذه الاختبارات، فهنالك اختبارات الصواب والخطأ، واختبارات المزاوجة، واختبارات ملء الفراغ، واختبارات الاختيار من متعدد (Multiple Choice Question) التي تُعدّ كما يشير كل من هالدينا وداوننج (Haladyna and Downing, 2004) أكثر أنواع الأسئلة الموضوعية استخدامًا في المؤسسات التعليمية؛ ولاسيما في المواقف التي تهدف لاختبار إعدادٍ كبيرة من المفحوصين، حيث يتطلب هذا النوع من الاختبارات من المستجيب أن يختار إجابة واحدة فقط من مجموعة من الإجابات المطروحة، ويتكون السؤال من قسمين رئيسين هما: متن السؤال أو الأرومة (Stem) الذي يحتوي على المشكلة، يليها مجموعة من البدائل أو المموهات بالإضافة للإجابة الصحيحة، وعمومًا يُعتبر هذا النمط من الاختبارات الموضوعية الأكثر استخدامًا عند بناء بنوك الأسئلة، أو إجراء الاختبارات التكيفية (Adapting Tests) أو غيرها من الاختبارات المختلفة في شتى المجالات؛ وذلك لما يتمتع به هذا النمط من الاختبارات الموضوعية من مرونة وقدرة على اختبار عددٍ كبيرٍ من المفحوصين، كما أنها تصلح لتقويم معظم نتاجات التعلم (Ebel and Frisbi, 1991; Merritt, 2006 ;Poham, 2003).

وعلى الرغم من الميزات التي تمتاز بها فقرات اختبار الاختيار من متعدد، إلا أنها تفتقرُ لإمكانية الكشف عن قدرة الطلبة على الإبداع أو الإتيان بأفكار عديدة ومتنوعة وأصيلة، ولا تستطيع فقراتها قياس قدرة الطلبة على التركيب والتقويم كعمليات عقلية عليا، كما أنه من الصعوبة بمكان الوقوف من خلالها على قدرة الطلبة على تنظيم المعلومات وربطها ببعضها البعض، إضافة إلى أنها تدفع المفحوصين للتركيز على تفاصيل صغيرة متناثرة وغير مترابطة من المحتوى الدراسي، حيث يُشير كل من شنايدر وايجن (Schneider, Egan and Julian, 2013) في هذا السياق إلى أنه حينما يختار الطالب بديلاً خاطئًا عند إجابته عن فقرات اختبار الاختيار من متعدد، فإن المفهوم الخطأ يتعزز لديه؛ وبهذا قد يكتسب الطالب معرفة غير صحيحة، وهذا ما يُسمى بحسب كل من شنايدر وايجن بـ "تأثير البديل الخطأ " (Negative Suggestion Effect)، ممّا يتطلب من المعلم مراجعة إجابات الاختبار مع الطلبة والوقوف على الأساس المعرفي لكل إجابة صحيحة، كما أن بنية أسئلة فقرات الاختيار من متعدد تؤدي إلى لجوء الطلبة للغش والتخمين العشوائي، ولعلّ لجوء الطالب للتخمين العشوائي عندما لا تتوافر لديه المعلومات التي تُمكنه من الوصول إلى الإجابة الصحيحة من أكبر المشكلات التي تواجه هذا النوع من الاختبارات، حيث يقوم الطالب بالاختيار العشوائي لأحد البدائل حتى لو طُلب منه أن يترك الفقرة بدون إجابة إذا لم تكن لديه معرفة بالإجابة الصحيحة، فبعض الطلبة يميلون إلى المجازفة أو المخاطرة بدرجات مختلفة حسب درجة التزامهم بالتعليمات، وحكمتهم في الإجابة، فأحيانًا قد توجد منبهات أو مفاتيح في الفقرة أو في البدائل تساعد الطالب على التوصل للإجابة الصحيحة، وبهذا لا يمكن للمصحح أن يقرر فيما إذا كانت الإجابة الصحيحة للطالب عن فقرةٍ ما تعكس قدرته الحقيقية على الإجابة وتمّكنه من المفهوم أو التعميم أو العلاقة المقاسة بالفقرة، وخاصة عندما يكون الهدف من الاختبار مثلاً تشخيص صعوبات التعلم أو الوصول الى درجة الإتقان لمهارة معينة(Annie and Chan, 2012).

ويُشير كلٌ من كبنجر وهولشر وريف (Kubinger, Holocher and Reif, 2010) إلى أن التخمين في الاختبارات الموضوعية ولاسيما الاختيار من مُتعدد، يؤدّي إلى غياب خاصية العدالة عن نتائج الاختبار؛ فمثلاً إذا ما تقدم مفحوصان من مستوى قدرة متوسطة لاداء اختبار موضوعي من نوع الاختيار من متعدد فإن هناك احتمالاً بأن يجتاز المفحوص الذي يلجأ للتخمين الاختبار أو يحصل على درجة أعلى مقارنة بالمفحوص الذي أجاب عن فقرات الاختبار بدون تخمين، وأيضا إذا كان هناك مفحوصان من مستوى قدرة متفاوتةً فقد ينجح الطالب ذو القدرة المنخفضة نتيجة لجوئه للتخمين، ويفشل الطالب ذو القدرة المرتفعة الذي أجاب عن فقرات الاختبار بناءً على معرفته الحقيقية، مما يؤدي إلى انتهاك أهم الخصائص الأساسية للاختبار وهي العدالة فتصبح نتائج الاختبار متحيزة.

فالتخمين العشوائي يعُدّ من المُهددات لكل من صدق وثبات اختبارات الاختيار من متعدد حيث يرى ميسك (Messick, 1995) أن لجوء الطلبة لعملية التخمين خلال إجابتهم عن فقرات اختبار الاختيار من متعدد يُضيف تباينًا لا يرتبط بالتباين في القدرة الحقيقية (أي ذلك التباين الذي يعكس قدرة الاختبار على قياس ما وُضع لقياسه)؛ ممّا يترتب على ذلك تخفيض صدق الاختبار بشكل ملموس؛ والذي يُقاس في إطار النظرية الكلاسيكية في القياس بنسبة التباين الحقيقي المرتبط بالسمة المقاسة من التباين الكلي، وإلى جانب ذلك فإن التخمين يُعّد نوعًا من الخطأ العشوائي في القياس؛ فهو أيضًا عامل مؤثر في دقة تقدير القدرة للمفحوصين، كما ويؤثر على بناء الاختبار حيث يتوقف قبول الفقرات أو رفضها على معاملي الصعوبة والتمييز اللذين يتأثران بدورهما بعامل التخمين، ويرى كل من بارهلل وبودسكيو واتالي (Bar-Hillel, Budescu and Attali, 2005) أن التخمين قد يؤثر أيضًا في ثبات الاختبار من ناحيتين، الأولى: إن ضيق الوقت في نهاية بعض الاختبارات قد يدفع الطالب إلى التخمين العشوائي السريع للإجابة عن باقي فقرات الاختبار، وهذا قد يؤدي إلى عدم اتساق إجابته عن هذه الفقرات المتبقية مع الإجابة عن الفقرات التي استغرق فيها وقتًا كافيًا للإجابة في بداية الاختبار؛ الأمرالذي يؤثر على ثبات ودرجة اتساق الإجابة عن الاختبار، أما الثانية فتتمثل في أنه عند تطبيق الاختبار مرة أخرى فمن المحتمل ألاّ يحصل المفحوص الذي يلجأ للتخمين العشوائي على الدرجة نفسها؛ لأنه قد يُغير إجابته عن فقرات الاختبار التي قام بتخمينها على أساسٍ غير منطقي(Camili, 2006;Mcmillan,2013) .

ونظرًا للآثار السلبية السابقة المترتبة على التخمين، فقد حظي باهتمامٍ بالغٍ من قبل العاملين في مجال القياس والتقويم سواء أكان في إطار النظرية الكلاسيكية للقياس أم في إطار نظرية الاستجابة للفقرة، فقد تضمّن الأدب النظري المتعلق بالقياس عددًا من الإجراءات الخاصة بالتعامل معه في إطار الأسئلة الموضوعية من نوع الاختيار من متعدد، حيث تم التوصل إلى عددٍ من النماذج الخاصة المنطقية والرياضية في نظرية الاستجابة للفقرة- والتي تم استعراضها وتناوُلها بشكل مُفصّل في إطار الفصل الثاني من هذه الدراسة- لمعالجة مشكلة التخمين كمحاولة للتخفيف من آثاره المترتبة على دقة تقدير معلمة القدرة للأفراد ولمعالم الفقرات.

إذ تُعدُّ طريقة عدد الإجابات الصحيحة وطريقة التصحيح ضد أثر التخمين من أكثر الطرق شيوعًا وألفةً بين المتعاملين مع الاختبارات الموضوعية/الاختيار من متعدد في إطار النظرية الكلاسيكية في القياس(Muijtjens,Mameren,Hoogenboom,Evers and Vleuten, 1999,p.267)؛ فقد يُطلب من المفحوص في تعليمات الاختبار عند تطبيق طريقة التصحيح ضد أثر التخمين(العقاب) ألاّ يلجأ إلى التخمين إذا شعر بأنه لا يمتلك المعرفة للإجابة عن الفقرة، وأن يقوم بترك الفقرة دون إجابة وكأنه يقول للمفحوص بأنك ستخسر إذا لم تلتزم بالتعليمات، وهناك طرق أخرى مختلفة

للتصحيح ضد أثر التخمين مثل: (المكافأة)، واستخدام البديل "لا أعرف الإجابة" بين البدائل، ومحاسبة الطلبة الذين يلجؤون للتخمين، وقد تم تناول هذه الطرق بالتفصيل في الفصل الثاني من هذه الدراسة (Camilli and Chiu, 2013).

عمومًا، يمكن القول إن طرق التصحيح لأثر التخمين في إطار النظرية الكلاسيكية في القياس تأخذ بعين الاعتبار المواقف الآتية عند إجابة المفحوص على فقرات الاختبار الموضوعي/الاختيار من متعدد اولاً: إن المفحوص يعرف الإجابة الصحيحة، وثانيًا: إن المفحوص لا يعرف الإجابة الصحيحة ويترك الفقرة دون إجابة، وثالثًاً: إن المفحوص لا يعرف الإجابة الصحيحة على الفقرة، ولكنه يُجيب عن الفقرة بشكل عشوائي، وقد لوحظ أنَّ بعض المطبقين لاختبارات الاختيار من متعدد يستخدمون طرق التصحيح التقليدية، ويستخرجون نتائج يبنون عليها قرارات وتوصيات متعددة، رغم أن جميع الدراسات في هذا الموضوع أثبتت عدم كفاية هذه الطرق لتصحيح الاختبارات الموضوعية/الاختيار من متعدد من حيث قدرتها على ضبط عامل التخمين (Sabbe,Valcke and Lesage, 2013).

أمّا في نظرية الاستجابة للفقرة؛ فقد ظهرت محاولات كثيرة للتغلب على مشكلة التخمين؛ تراوحت بين حذف تلك الفقرات التي يظهر من نمط إجابة المفحوص عليها أنه استخدم التخمين للوصول إلى الإجابة الصحيحة (Waller, 1989)، إلى النماذج المُعدّلة (Modified IRT Models)؛ حيث يتم إضافة قيمة ثابتة غير صفرية بدلاً من معلمة التخمين للنماذج الثنائية والأحادية تساوي احتمالية الإجابة عن الفقرة بالتخمين بشكل صحيح ($\frac{1}{A}$) حيث (A) تساوي عدد بدائل الفقرة (Barnes and Wise , 1991)، ثم جاءت النماذج المختلطة (Mixture Model) المقترحة من قبل يامتو (Yammato, 1995) والتي تقوم بالدمج في صيغتها الرياضية بين النماذج اللوجستية ومعاملات رياضية إضافية، فإذا لجأ المفحوص للتخمين خلال إجابته عن فقرات الاختبار الموضوعي/الاختيار من متعدد فإن النموذج المختلط يعمل، وإذا أجاب المفحوص عن الفقرة بدون تخمين فإن جزءًا من النموذج المختلط يعمل، وهذا ما سيتم توضيحه خلال استعراض الأدب النظري في الفصل الثاني من هذه الدراسة.

من هنا جاءت الدراسة الحالية إستكمالاً للجهود التي بُذلت في إطار كل من الأدب التجريبي والنظري الذي تناول موضوع التخمين ومعالجته أو التخفيف من آثاره في مجال القياس النفسي والتربوي، بهدف مقارنة فاعلية أربعة نماذج لمعالجة التخمين في الأسئلة الموضوعية/الاختيار من متعدد في إطار النموذج اللوجستي ثلاثي المعلمة وأثرها على تقدير معلمة القدرة ومعالم الفقرات

وذلك باستخدام بيانات مولدة وذلك للوقوف على كيفية عمل هذه الإجراءات في ظل مواقف نموذجية مضبوطة قد لا تتوفر بالواقع التجريبي بالشكل الذي قد توفره البيانات المولدة.

إذ جرى اختبار فعالية نماذج وضعت من قبل كل من كاو وستوكس (Cao and Stokes, 2008) تستند إلى طريقة التقدير وفق نظرية بيزن (Baysian)؛ حيث راعت هذه النماذج أنماطاً مختلفة من سلوك التخمين التي قد يقوم بها المفحوص، وهذه النماذج هي: نموذج عتبة التخمين والنموذج القائم على الصعوبة، كما تم استخدام نموذجين آخرين هما: النموذج اللوجستي الثلاثي المعلمة، والنموذج اللوجستي الثنائي المعلمة المعدّل إلى ثلاثي من خلال تثبيت معلمة التخمين عند القيمة (٠,٢٠).

هدف الدراسة وأسئلتها

هدفت هذه الدراســـة إلى الوقوف على فاعلية أربع طرق ونماذج لمعالجة التخمين في الأسـئلة الموضـوعية /الاختيار من متعدد ضـمن إطار النموذج اللوجسـتي الثلاثي المعلمة وأثرها على تقدير معلمـة القدرة ومعـالم الفقرة بـإختلاف متغير طول الاختبـار (٢٥ فقرة، و ٥٠ فقرة) وعدد المفحوصـــين (٥٠٠ مفحوص و ١٥٠٠ مفحوص). إذ هدفت هذه الدراســـة إلى الإجابة عن الأسئلة الآتية:

السؤال الاول

هل تختلف دقة تقدير معلمة القدرة باختلاف النموذج المســتخدم لمعالجة التخمين في إطار النموذج اللوجســتي الثلاثي المعلمة تبعًا لمتغيري طول الاختبار (٢٥ فقرة و٥٠ فقرة) وعدد المفحوصــين (٥٠٠ مفحوص، و١٥٠٠ مفحوص)؟.

السؤال الثاني

هل تختلف دقة تقدير معالم الفقرات باختلاف النموذج المستخدم لمعالجة التخمين في إطار النموذج الثلاثي المعلمة تبعًا لمتغيري: طول الاختبار (٢٥ فقرة و٥٠ فقرة)، وعدد المفحوصـــين (٥٠٠ مفحوص و١٥٠٠ مفحوص)؟.

السؤال الثالث

هل يختلف مؤشـــر الكفاءة النســـبية لتقدير معالم الفقرات ومعلمة القدرة للمفحوصـــين بإختلاف متغيرات الدراسة (النموذج المستخدم لمعالجة التخمين وعدد الفقرات وحجم العينة)؟

أهمية الدراسة

تكمن أهمية هذه الدراسة في الجوانب الآتية:

أ- محاولتها توفير معلومات حول فاعلية طرق معالجة التخمين التي تؤدي إلى دقة تقدير معالم الفقرات، ومعلمة القدرة للمفحوصين في إطار النموذج اللوجستي الثلاثي المعلمة، عندما يتم استخدام أربعة نماذج لمعالجة التخمين، وذلك باستخدام محكات واحدة لمحاكمة الفاعلية؛ ممّا يُوفر أرضية مشتركة للمقارنة بين هذه النماذج؛ وممّا يُسهم كذلك في تحديد النموذج الأكثر فاعلية للتخفيف من آثار مشكلة التخمين على تقدير معالم الفقرات ومعلمة القدرة للمفحوصين، تحت الظروف المختلفة لعدد الفقرات وحجم عينة المفحوصين، الأمرالذي يُسهم في توجيه المستخدمين لهذه النماذج والمهتمين بمعالجة التخمين إلى توظيف النموذج الأنسب للتعامل مع التخمين وبالتالي توفير تقديرات أكثر دقة لمعالم الفقرات ومعلمة القدرة للمفحوصين في الاختبارات الموضوعية/الاختيار من متعدد.

ب- كما يمكن أن توفر هذه الدراسة المعلومات الكافية لمستخدمي الاختبارات الموضوعية/الاختيار من متعدد سواء في حقل التدريس أو في غيره من الحقول حول مزايا وإيجابيات هذا النمط من الاختبارات ونقاط الضعف، وكيفية التغلب عليها وحول أنسب الطرق لتصحيح الفقرات.

ج- إثراء الدراسات المرتبطة بموضوع التخمين في الاختبارات الموضوعية/الاختيار من متعدد، من خلال الكشف عن أثر التخمين على تلك الاختبارات.

د- توظيف الطريقة البييزية(Bayesian) في تقدير معالم الفقرات ومعلمة القدرة للمفحوصين والتي تتميز عن طريقة الأرجحية العظمى؛ بعدم إعطاء أخطاء عندما تكون إجابة المفحوص جميعها خطأ أو جميعها صحيحة (Hambelton, 1989)، إذ أن هناك قلة في الدراسات التي استخدمت الطريقة البييزية في سعيها للتغلب على مشكلة التخمين.

هـ- استخدامها برمجية واحدة فقط لتقدير معلمة القدرة ومعالم الفقرات لجميع النماذج المستخدمة في هذه الدراسة، ممّا يعني وجود أرضية واحدة للمقارنة بين هذه النماذج، حيث يتصف هذا البرنامج بالدقة والشمولية والقدرة على التعامل مع كافة أنواع النماذج اللوجستية أحادية البعد أو متعددة الأبعاد والنماذج المختلطة (Mixture Model)، الأمر الذي سيتيح للباحثين إمكانية التعامل مع هذا البرنامج والتعرف على إمكانياته ومزاياه البحثية، ووضع توصيات خاصة لبناء أدوات

القياس التي تتضـــمن فقرات اختيار من متعدد بحيث تكون النتائج عليها إلى حد كبير متحررة من قضية التخمين.

و-إلقاء الضـــوء على النماذج المختلطة باعتبارها إحدى الطرق المطروحة للتعامل مع مشـــكلة التخمين.

مصطلحات الدراسة وتعريفاتها الاجرائية

سيكون للمفاهيم والمتغيرات المعاني الآتية في إطار هذه الدراسة:

- **معالم الفقرات** (Item Parameter) :

تم استخدام مفهوم معالم الفقرات في إطار هذه الدراسة لتشير إلى كل من : معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين.

- **معلمة صعوبة الفقرة:**

تم النظر لها في هذه الدراسة على أنها النقطة على متصل القدرة والتي يكون احتمال إجابة الفقرة عندها إجابة صحيحة مساويًا لـ (C_i+1) مقسوما على ٢ ، حيث (C_i)هي معلمة التخمين، وجرى تقديره باستخدام برمجية (Winbug).

- **معلمة تمييز الفقرة:**

جرى النظر لمعلمة تمييز الفقرة في هذه الدراسة على أنها ميل منحنى خصائص الفقرة الذي يقابل النقطة التي تكون عندها معلمة القدرة تساوي معلمة صعوبة الفقرة، وجرى تقديرها باستخدام برنامج (Winbug).

- **معلمة التخمين الفقرة:**

تم النظر إلى معلمة التخمين في هذه الدراسة على أنها احتمال حصول المفحوص ذي القدرة المتدنية على الإجابة الصحيحة بالصدفة، وجرى تقديرها باستخدام برنامج (Winbug).

- **معلمة القدرة للمفحوصين** Ability Parameter:

تم النظر لمعلمة القدرة في هذه الدراسة على أنه مقدار ما يمتلكه الفرد من السمة التي يتم قياسها على متصل السمة والذي تتراوح قيمتها نظريًا بين (-∞ و+∞)، وعمليًا بين (-3 و 3)، وجرى تقديرها من خلال برنامج (Winbug).

- **الطريقة البييزية في التقدير** Bayesian Estimation:

تم في هذه الدراسة استخدام الطريقة البييزية لتقدير معالم الفقرة: (التمييز والصعوبة والتخمين) ومعلمة القدرة إذ تقوم هذه الطريقة على أساس افتراض أن هناك توزيعًا احتماليًا مسبقًا قد يسهم في تحقيق دقة أكبر في عملية التقدير.

-الخطأ المعياري للتقدير Standard Error of Estimate:

تم النظر إلى الخطأ المعياري للتقدير بأنه ذلك المقياس للتشتت المُرتبط بقيم (θ) المقدرة لعدد من المفحوصين حول قيم (θ) الحقيقية لهم.

- دالة معلومات الاختبار Test Information Function:

ونظر لدالة المعلومات للاختبار في هذه الدراسة على أنها كمية المعلومات التي يمكن من خلالها أن تُميز الفقرة بين مستويات القدرة، ويتم إيجادها من خلال جمع دوال المعلومات للفقرات المكونة للاختبار عند مستوى قدرة(θ).

-النموذج اللوجستي ثلاثي المعلمة Three-Parameter Logistic Model (3PL):

ونُظر له في هذه الدراسة على أنه النموذج الذي يتحدد به الأداء على الفقرة بناءً على جميع معالم الفقرة: معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين.

- نموذج عتبة التخمين IRT Threshold Guessing Model:

جرى النظر له في هذه الدراسة كأحد أشكال النماذج المختلطة التي تدمج بين النموذج اللوجستي الثنائي المعلمة ومعاملات رياضية إضافية، فإذا سلك المفحوص سلوك التخمين خلال إجابته عن فقرة الاختبار يعمل النموذج المختلط، وإذا لم يُخمّن يوظف النموذج اللوجستي الثنائي المعلمة، ويحتوي النموذج على معلمة تُحدد لكل فرد العتبة التي يُخمن عندها، ويقوم على افتراض أن المفحوص يُجيب عن فقرات الاختبار إلى حد معين، ثم يستخدم التخمين العشوائي للإجابة عن باقي فقرات الاختبار.

- نموذج التخمين القائم على الصعوبة IRT Difficulty-Based Guessing Model :

جرى النظر له في هذه الدراسة على أنه شكل من أشكال النماذج المختلطة التي تدمج بين النموذج اللوجستي الثنائي المعلمة ومعاملات رياضية إضافية، ويفترض أن المفحوص يجيب عن الفقرات السهلة أولاً، ثم يستخدم التخمين للإجابة عن باقي الفقرات التي مستوى صعوبتها أعلى من مستوى قدرته.

- النموذج الثنائي المعدل:

ويُنظر للنموذج الثنائي المعدل في هذه الدراسة على أنه النموذج اللوجستي الثنائي المعلمة المُعّدل إلى ثلاثي من خلال إضافة قيمة ثابتة عوضًا عن معلمة التخمين تبلغ (٠,٢٠).

- الكفاءة النسبية Relative Efficiency :

جرى تقدير الكفاءة النسبية باختلاف متغيرات الدراسة من: حجم عينة وعدد فقرات الاختبار والنموذج المستخدم في الدراسة لمعالجة التخمين من خلال قسمة تباين معالم الفقرة ومعلمة القدرة

التي تم تقديرها من خلال نموذج معين على تباين نظيراتها المقدرة من خلال نموذج آخر فإذا زاد ناتج القسمة عن الواحد الصحيح تكون الكفاءة النسبية لصالح النموذج الأول وإذا كان الناتج أقل من الواحد الصحيح تكون الكفاءة النسبية لصالح النموذج الثاني.

- **دالة معلومات الفقرة** Item Information Function :

ونُظر لدالة المعلومات للفقرة في هذه الدراسة على أنها اقتران رياضي يُعبر عن كمية المعلومات المتمثلة في تمييز الفقرة بين مستويات القدرة المختلفة.

- **دقة التقدير** Accuracy of estimating:

جرى الحكم على دقة التقدير في إطار هذه الدراسة باستخدام الجذر التربيعي لمتوسط مربع الخطأ للتقدير، ومتوسطات الخطأ المعياري للتقدير، والكفاءة النسبية، ومؤشر التحيز.

-**اللوجت**: وحدة قياس كل من معلمة الصعوبة ومعلمة القدرة للمفحوصين.

- **الجذر التربيعي لمتوسط مربع الخطأ** (RMSE) Root Mean Square Error **:**

تم النظر إلى الجذر التربيعي لمتوسط مربعات الخطأ على أنه مؤشر لدقة تقدير معالم الفقرات ومعلمة القدرة، حيث يتم إيجاد الجذر التربيعي لمتوسط مربعات الفروق بين القيم المولدة(الحقيقية) والقيم المقدرة.

التحيز (Bias)**:**

تم النظر للتحيز في هذه الدراسة بأنه مؤشر يُعبر عن الدرجة التي تكون فيها القيم المقدرة لمعالم الفقرات وفق النموذج المستخدم لمعالجة التخمين في هذه الدراسة قريبة من القيم الحقيقية (المولدة).

محددات الدراسة

تتحدد نتائج هذه الدراسة بالمحددات الآتية:

أ- البيانات التي تم استخدامها في إطار هذه الدراسة: حيث جرى توليد بيانات مولدة (محاكاة) باستخدام برنامج (WinGen3) بدلاً من بيانات حقيقية جُمعت من مفحوصين حقيقيين.

ب- النماذج الخاصة بتقدير التخمين التي وظُفت في إطار هذه الدراسة هي:

النموذج اللوجستي الثلاثي المعلمة، و النموذج اللوجستي الثنائي المعدل إلى ثلاثي بإضافة قيمة ثابتة كمعلمة تخمين تساوي (٠,٢٠)، ونموذج عتبة التخمين، والنموذج القائم على الصعوبة.

ج- تقدير معالم الفقرات ومعلمة القدرة للمفحوصين باستخدام برنامج: (Winbug)

(Windows Bayesian Using Gibbis Sampling) وباستخدام التقدير البييزي.

د- البيانات المّولدة كانت لاختبارات موضوعية/اختيار من متعدد بافتراض أربعة بدائل للفقرة.

الفصل الثاني

الإطار النظري والدراسات السابقة

الفصل الثاني

الإطار النظري والدراسات السابقة

مقدمة

يتناول هذا الفصـــل أبرز الجهود التي بُذلت في إطار الأدب النظري والتجريبي لمعالجة مشـكلة التخمين في الأسـئلة الموضـوعية/ الاختيار من متعدد سـواء أكاَن ذلك في إطار النظرية الكلاسـيكية في القياس أم في إطار نظرية الاسـتجابة للفقرة، كما يتناول هذا الفصـل أبرز النماذج الرياضية التي وُضعت للتعامل مع مشكلة التخمين في الأسئلة الموضوعية/الاختيار من متعدد في إطار النموذج اللوجستي ثلاثي المعلمة والمؤشـرات التي وُضـعت للحكم على فاعليتها، كما جرى اسـتعراض أبرز الجهود والدراسـات السـابقة التي أجريت حول موضـوع الدراسـة، للوقوف على فاعلية الطرق والنماذج الخاصة بمعالجة التخمين في الأسئلة الموضوعية/الاختيار من متعدد .

مفهوم التخمين

يفترض العاملون في مجال قياس التحصـيل أنّ المفحوص عندما يجيب عن فقرات الاختبارات الموضـوعية/الاختيار من متعدد فإنه سـيجيب عن كل فقرة وفقًا لقدرته الحقيقية؛ بمعنى أنه سـوف يختار البديل الذي يحتوي الإجابة الصـــحيحة إذا كان يمتلك المعرفة حول تلك الإجابة، ولكن هذا الافتراض لا يعكس دائمًا بدقة ما يحصـــل على أرض الواقع؛ ولا ســيّما في الاختبارات الموضـــوعية/الاختيار من متعدد حيث يُمكن تصـــنيف إجابات المفحوصـــين في ثلاث فئات: (Mehrens and Lehman,1991)

١-إجابات تعكس القدرة الحقيقية (سواء صواب أو خطأ).

٢-إجابات صحيحة ناتجة عن تخمين محظوظ.

٣-إجابات خطأ بالرغم من امتلاك الطالب أو المفحوص للإجابة الصحيحة.

إذ لا توجد مشكلة لدى العاملين في مجال القياس والتقويم مع المفحوصين ضمن الفئة الأولى ولكن تكمن المشـكلة في التقليل من عدد الأفراد الذين هم في الفئة الثانية والفئة الثالثة على حسـاب زيادة عدد الأفراد في الفئة الأولى(Liao,2012).

يُعدّ التخمين أسـلوبًا يلجأ إليه المفحوص عند إجابته عن فقرات الاختبار الموضوعي/الاختيار من متعدد عندما لا تكون لديه المعرفة التامة بالإجابة الصحيحة للفقرة، أو عندما يكون لديه معرفة جزئية تجعله قادرًا على حصر الإجابة الصحيحة بين عددٍ محدودٍ من البدائل للفقرة، فيختار أحدها، أو قد لا تكون لديه مثل هذه المعرفة الجزئية فيختار عشوائيًا من بين جميع البدائل الموضوعة للفقرة كإجابة صحيحة (Reif, Ertl, Stefana , Kubinger and Martina, 2010).

كما ويرى روجرز (Rogers,1999) أن التخمين يُقسم إلى ثلاثة أنواع هي:

أ- التخمين العشوائي:

ويُنظر إليه على أنه التخمين الذي يتم دون ترجيح لأحد البدائل على الآخر، وهنا سيكون احتمال الحصول على الاجابة الصحيحة مساويًا لــ ($\frac{١}{\text{عدد البدائل}}$)، إذ يحدث هذا النوع من التخمين حينما لا يكون لدى المفحوص أي فكرة عن الإجابة الصحيحة، وبالتالي يختار عشوائيًا من بدائل الفقرة، فيكون اختياره للبديل الذي يتضمن الإجابة الصحيحة بالحظ، ومن هنا فإَنّ التخمين العشوائي لا يُعطي أية معلومات عن توفر القدرة لدى المفحوص.

ب- التخمين المبنيّ على المعرفة الجزئية لدى المفحوص:

وفي هذا النوع من التخمين يقوم المفحوص باستبعاد بعض البدائل غير الصحيحة، ويحصر الإجابة بين عدد أقل من البدائل.

ج-التخمين المبني على التلميحات (Cued Guessing):

عندما يختار الطالب البديل بناءً على دلائل ومحكات موجودة في بُنية الفقرة تقوده للإجابة الصحيحة، وقد تكون هذه الدلائل صيغة السؤال مثل بعض الكلمات في البدائل المختلفة للفقرة أو للفقرات الأخرى التي يستدلّ منها المفحوص على الإجابة الصحيحة أو الإجابة الخطأ؛ وبالتالي يتمكّن من استبعاد بدائل خطأ وحصر الإجابة الصحيحة بين عدد أقل من البدائل ثم يلجأ للتخمين العشوائي بين هذا العدد الأقل من البدائل، وعندها يصبح احتمال الحصول على الإجابة الصحيحة كبيرًا.

لقد حظي التخمين المبني على معرفة جزئية لدى المفحوصين باهتمام أقل من قبل المختصين والتربويين؛ بل إن كلّ من باشمان و بالمر(Bachman and Palmer,1996) يقترحان تشجيع المفحوصين على ممارسة هذا النوع من التخمين وذلك بتوظيفهم للمعرفة الجزئية عن الموضوع الذي تقيسه فقرة الاختبار عند الاجابة عن فقرات الاختبار الموضوعي/الاختيار من متعدد من خلال حذف واستبعاد البدائل التي لا تبدو قريبة من الجواب الصحيح، ويضيفان أنه بدلاً من اللجوء إلى معادلات التصحيح لأثر التخمين يجب التخلص من الأسباب التي تقود المفحوصين للتخمين من خلال :

١-إعطاء المفحوصين زمنًا كافيًا للإجابة عن فقرات الاختبار؛ لأنّ قصر الزمن المخصص للإجابة يقود المفحوصين لاستخدام التخمين العشوائي في اختيار بدائل فقرات الاختبار.

٢-اختيار الفقرات التي تقترب من مستوى قدرة المفحوصين.

وقد أشار النبهان (Alnabhan,2002) إلى أن التخمين العشوائي يُعتبر مصدرًا لكل من: الخطأ العشوائي والخطأ المنتظم في الاختبار، فالخطأ المنتظم يؤثر على صدق الاختبار ولا يؤثر على ثبات الاختبار، في حين يؤثر الخطأ العشوائي على كل من صدق الاختبار وثباته، كما يرى كل من هالدينا وداوننغ (Haladyna and Downing,2004) أن الخطأ المنتظم في الاختبار هو مصدر التباين غير المرتبط بالسمة المقاسة.

كما اعتبر كل من وايز وكونج (Wise and Kong ,2005) أيضًا أن التخمين مكونُ من مكونات خطأ القياس المنتظم وهو يقود إلى إحداث فرق بين القدرة الحقيقية للمفحوص وتلك المُقدرة، كما قد يقود هذا النوع من التخمين إلى مشاكل تتمثل في عدم مناسبة نموذج نظرية الاستجابة للفقرة للبيانات (Bovaired ,2002; Messick, 1995).

وأشار وايز وكونج إلى أن بعض المفحوصين يميلون أحيانًا في ظلّ بعض الظروف إلى تخمين الإجابة الصحيحة بسرعة عند أدائهم على اختبارات الاختيار من متعدد، وهذا ما يُسمّى بالتخمين السريع (Rapid Guessing)، ويُعزى سلوك التخمين السريع إلى عدم كفاية زمن الامتحان، أو انعدام الدافعية لدى المفحوص للإجابة لِعدم أهمية الاختبار بالنسبة له وعدم ترتب أي نتائج على درجاته.

كذلك يُعدّ التخمين سمة شخصية للفرد، فنزعة الأفراد إلى التخمين تختلف من فرد إلى آخر، وهو يُكتسب من قبل الفرد، حيث يتعلم كيف ومتى يستخدم التخمين، فالبيئة والتعليم والثقافة والدافعية عوامل كلها تؤثر في انتهاج المفحوص لسلوك التخمين من عدمه (Xie, 2002).

وفي هذا السياق يضيف شوبن (Choppin,1975) أن النزعة إلى التخمين هي نتيجة تفاعل بين المفحوص والفقرة، ولا تبدو سِمة منتظمة عبر الأفراد، وقد أثبتت نتائج دراسات أجريت على اختبارات العلوم والرياضيات العالمي (TIMSS) Trends International Study of Mathematics and Science وغيرها من الاختبارات الدولية أن بعض الطلاب في دول معينة يلجأون إلى سلوك التخمين أكثر من غيرهم من الطلاب في دول أخرى؛ فمثلاً وجُد أن الطلاب في الولايات المتحدة وبريطانيا يُخمنون أكثر من الطلاب في الدول الأوربية خاصة فرنسا والسويد؛ بينما الطلاب في الدانمارك ودول شرق آسيا يتركون الفقرة دون إجابة ولا يُخمنون عندما لا يمتلكون المعرفة التامة للإجابة عن الفقرة (Adams and Wu,2000 ; Bolger, Kellaghan,1990).

يلجأ بعض التربويين للتغلب على مشكلة التخمين وتحسين ثبات الاختبارات الموضوعية/الاختيار من متعدد إلى زيادة عدد الفقرات المكوّنة للاختبار وزيادة عدد البدائل للفقرة الواحدة فيه؛ نظرًا لأن نسبة التخمين العشوائي على فقرات اختبارات الاختيار من متعدد تقل بزيادة عدد البدائل فهي حوالي (٢٠%) في حالة وجود خمسة بدائل و (٢٥%) في حالة وجود أربعة بدائل و (٣٣%) في حالة وجود ثلاثة بدائل للفقرة، إلا أن هذا الإجراء يؤخذ عليه أنه يقود إلى زيادة الوقت الذي يحتاجه المفحوص للإجابة عن مثل هذه الفقرات، كما أن زيادة عدد البدائل للفقرة تُقلل من إمكانية وضع مموهات أو بدائل جاذبة للمفحوصين، علمّاً أن قدرة المموهات على جذب المفحوصين والتجانس بينها يعتبران من الخصائص الأساسية التي يجب أن يتمتع بها هذا النوع من الفقرات، وبالتالي فإنه لا فائدة من وضع عددٍ كبيرٍ من البدائل للفقرات دون تمتعها بمثل هذه الخصائص(Haladyna, 2004).

وفي هذا السياق يُشير دي بوك (San Martin, Del Pino and De Boeck, 2006) إلى وجود سببين للاعتقاد أن التخمين مرتبط بقدرة الفرد:

أ- الأول: إن الأفراد عندما يواجهون فقرة لديهم معرفة جزئية حولها يحاولون اللجوء إلى تخمين الإجابة الصحيحة لها، خاصة ذوي القدرة المنخفضة فكلما زادت صعوبة الفقرة زاد سلوك التخمين من قبل المفحوص، وكلما سهلت الفقرة قلّ سلوك التخمين خاصة للأفراد ذوي القدرة المرتفعة.

ب- الثاني: يتمثّل في أنّ معلمة التخمين في النموذج اللوجستي الثلاثي المعلمة في نظرية الاستجابة للفقرة تزيد أحيانًا قيمتها عن ($\frac{1}{N}$) حيث (N) هي عدد البدائل للفقرة والتي تعادل احتمالية الإجابة الصحيحة بشكل عشوائي، والتفسير المعقول لذلك أنّ بعض المُمتحنين يُمكنهم حذف بديل أو أكثر ثم يقومون بالتخمين بين باقي البدائل، وبالتالي فإن المعرفة الجزئية تعكسها معلمة التخمين؛ ولذا فهي مرتبطة بالقدرة(San Martin and etl, 2006) .

كما أن عزل الدرجات المتحققة للمفحوص كنتيجة لجوئه للتخمين(العشوائي، أو المنطقي) يُعد ذا أهمية خاصة عند تقدير معلمة القدرة للمفحوص أو تحصيله الأكاديمي واتخاذ القرارات حوله، نظرًا لأنّ التخمين يزيد من قيمة الخطأ المقاس، وكذلك يزيد من قيمة التباين غير المرتبط بالسمة المقاسة وعندها قد ينعكس ذلك جوهريًا على تشبعات (Loading) فقرات الاختبار بهذا العامل، وبالتالي ينتج عدم تحقق افتراض أحادية البعد (Uni-Dimensionality) الذي تقوم نماذج نظرية الاستجابة للفقرة عليه، إذ تفترض جميعها أن الدقة في تقدير معالم الفقرة ومعلمة القدرة التي تقدمها البرامج الإحصائية القائمة على هذه النماذج مبنية على تحقق افتراض أحادية البعد.

طرق معالجة التخمين في إطار النظرية الكلاسيكية في القياس

حظي موضوع معالجة أثر التخمين في فقرات الاختيار من متعدد في الاختبارات الموضوعية باهتمامٍ متواصلٍ من علماء القياس النفسي والتربوي في النظرية الكلاسيكية في القياس إذ جرى بحثها للتعرف على مدى تأثيرها سواء على تقدير أداء المفحوصين أو على تقدير الخصائص السيكومترية لفقرات الاختبارات وصدق وثبات الدرجات عليها، إذ جرى النظر إلى الأداء على الفقرات الموضوعية/الاختيار من متعدد من منظور أن المفحوصين عند أدائهم عليها يُقسمون إلى مجموعتين: مجموعة المفحوصين الذين يجيبون إجابة صحيحة عن الفقرة ؛لأنهم اختاروا الإجابة الصحيحة عن معرفة تامة أو حصلوا على الإجابة الصحيحة بواسطة التخمين(سواء كأن لديهم معرفة جزئية أو لم يكن) والمجموعة الثانية من المفحوصين أولئك المفحوصون الذين أجابوا عن الفقرة خطأ لأنهم إما يمتلكون معلومات غير صحيحة تمامًا أو لديهم معلومات غير صحيحة جزئيًا فيختارون أحد البدائل خطأ باستخدام التخمين، وقد يكون بعضهم قد أخطأ في اختيار البديل الصحيح لسبب ما مع معرفتهم له، وهكذا فإن الإجابة عن الفقرة قد تكون خطأ مع أنّها لم تكن نتيجة تخمين عشوائي(Alnabhan ,2002).

وبناءً على ذلك يمكن القول إن العاملين في النظرية الكلاسيكية غير قادرين على تدريج المفحوصين على سلم أو متصل وفق ما لديهم من معرفة، يمتد من المعرفة الكاملة بالمعلومات

المقاســـة بالســـؤال أو الفقرة إلى عدم المعرفة التامة بالمعلومات المقاســـة بالفقرة، كما أنهم غير قادرين على ضــبط أثر التخمين في الاختبارات الموضــوعية/الاختيار من متعدد؛ ممّا فتح المجال أمام المهتمين بالقياس والتقويم بالبحث عن طرق تأخذ بعين الاعتبار مســـتويات المعرفة عند المفحوصـــين، وتكون قادرة على التقليل من أثر التخمين، الأمر الذي ترتب عليه بروز عدد من المحاولات والاقتراحات التي طورت من أجل التعامل مع التخمين ومعالجته، وفيما يلي استعراضٌ لأبرز الطرق التي طُورّت في إطار النظرية الكلاسيكية في القياس لمعالجة وضبط أثر التخمين في إطار الاختبارات الموضوعية /الاختيار من متعدد.

أولاً: طريقة معادلة التصحيح لأثر التخمين:

تُعتبر هذه الطريقة من أكثر الطرق شـــيوعًا، وتقوم على معاقبة المفحوص جراء لجوئه للتخمين خلال إجابته على فقرات الاختبارات الموضـــوعية/الاختيار من متعدد عن طريق حذف الدرجات التي يُقدر أنهم قد حصـــلوا عليها من خلال التخمين، إذ تفترض هذه الطريقة أن كل اســـتجابة خطأ هي نتيجة تخمين خطأ قام به المفحوص، حيث لا تأخذ هذه الطريقة بعين الاعتبار المعلومات الجزئية التي ربما يكون قد استند إليها المفحوص في إجابته عن فقرات الاختبار، وتبعًا لهذه الطريقة تقدّر علامة المفحوص على الاختبار بعدد الإجابات الصـــحيحة مطروحًا منه (عدد الإجابات الخطأ) مقسومًا على (عدد بدائل الفقرة-١) ويُعبر عن ذلك كما في المعادلة الآتية:

$$S = R - \frac{W}{(A-1)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots [1]$$

(Mehrens and Lehmann, 1991, p463)

حيث إن:

S : الدرجة المصحّحة.

R : عدد الإجابات الصحيحة.

W :عدد الإجابات الخطأ ، A : عدد البدائل للفقرة.

وتبعًا لســـابي وفالكي ولســـج (Sabbe, Valcke and Lesage,2013) هذه الطريقة لا تُشـــجع على التخمين العشـوائي الذي يُضـخم درجات المفحوصـين، كما إنها تعطي تقديرًا غير متحيز للمعرفة الحقيقية التي يمتلكها المفحوص بناءً على أدائه في الاختبار، كما أنها تزيد من صـــدق الاختبار وثباته مقارنة مع طريقة التصحيح القائمة على عدد الاجابات الصحيحة.

بينما يُشـــير كل من مهرن ولهمن (Mehrens and Lehmann,1991) إلى أن اســـتخدام المعادلة الســـابقة يميل لمُعاقبة المفحوصـــين ذوي المعرفة الجزئية أكثر من ذوي المعلومات غير الصـحيحة، وفي هذا السياق يقترح فراي (Fray ,1988) لمعالجة ميل هذه الطريقة لمعاقبة التخمين المترتب على المعرفة الجزئية أن يعُطى المفحوص تعليمات تُفيد أن عليه أن يجيب عن كل فقرة يشعر أن لديه عنها معرفة جزئية بها ترشده لاختيار البديل الصحيح، أو تمكّنه من استبعاد أيٍ من البدائل التي يعتقد جازمًا أنها غير صـحيحة، ثم يقوم بالتخمين من بين ما تبقى من البدائل، و يجب أن تكون تعليمات الاختبار واضـــحة لإعلام المفحوص وإقناعه باتباع هذه الاســـتراتيجية عند أخذ الاختبار، وعندها يرى فراي أن اســتخدام معادلة التصـــحيح الســـابقة في ظل هذه التعليمات يُمَكن المفحوص من اسـتغلال المعرفة الجزئية لديه، إذ إنه سـيحذف فقط الفقرات التي يعتقد أن إجاباته عنها ناتجة عن التخمين العشوائي.

ويشير بلس (Blis,1980) أن الاقتراح السابق لـــــ فراي يقود إلى افتراض أن الفرق بين ورقة اختبار صُـحّحت وفق تعليمات معادلة التصـــحيح لأثر التخمين، ونفس الورقة إذا صُـحّحت تحت تعليمات إجابة جميع الفقرات، هو فقط في أن الفقرات المحذوفة إن وجدت في الحالة الأولى، تكون قد استبدلت بالتخمين العشـوائي في الأخيرة، وهذا يعني أن المفحوص إما أنه يعرف الإجابة أو يُخمّن بشكل عشوائي.

من جهة أخرى يرى المعارضـــون لاســتخدام معادلة التصـــحيح لأثر التخمين أن ترتيب المفحوصـــين بالنسـبة لبعضـهم البعض اعتمادًا على درجات الاختبار لا يتغير سـواء اسـتخدمت الدرجات الكلية دون تصـــحيحها من أثر التخمين أو بعد إجراء هذا التصـــحيح، كما يرون أنها لا تناسـب تصـحيح أثر التخمين المبني على معلومات جزئية، وكذلك يرون أنها تسـتند إلى افتراض خـاطئ مؤداه أن درجـات الاختبـار بعـد تصـــحيحهـا من أثر التخمين تكون هي الـدرجـات الحقيقية (Ebel and Frisbi,1991)، كما أن هناك جدلاً حول مقدار القيمة التي سـيتم حسـمها من درجة المفحوص عندما يجيب إجابة غير صـــحيحة عن فقرة اختبار الاختيار من متعدد فالقيمة التالية: ($\frac{1}{n-1}$) ، (حيث n تمثل عدد بدائل الفقرة) تعتبر غير كافية لكي تحد من اســـتخدام المفحوصين لسلوك التخمين العشوائي خلال إجابتهم عن فقرات الاختبار الموضوعي/الاختيار من متعدد (Espinosa and Gardeazabal,2010; Frary,1988).

ثانيا: طريقة معادلة التصحيح لأثر التخمين (المكافأة)

تم اقتراح هذه الطريقة من قبل كل من راولي وتروب (Rowley and Traub, 1977) حيث تقوم هذه الطريقة على مكافأة المفحوص لعدم انتهاجه سلوك التخمين خلال إجابته عن فقرات اختبار الاختيار من متعدد، من خلال إضافة قيمة لدرجته الكلية تعادل عدد الفقرات التي قام بحذفها أو عدم الإجابة عنها مقسومًا على عدد البدائل لكل فقرة، وعدم اتخاذ أي إجراء فيما يتعلق بالإجابات الخطأ وتفترض هذه الطريقة أن المستجيب يتبنى ثلاثة مواقف محتملة للاستجابة عن الفقرة في الاختبارات الموضوعية/الاختيار من متعدد:

١- يعرف المستجيب الإجابة الصحيحة ويقوم باختيارها.

٢- يحذف المستجيب الفقرة أو لا يُجب عنها.

٣- يُخمّن المستجيب بشكل عشوائي عند الإجابة عن الفقرة.

(Camilli and Chiu,2013; Fray,1988; Kruz,1999)

وبالتالي فإن الصيغة الأساسية لمعادلة تصحيح العلامات من أثر التخمين تكون كما يلي:

$$R_C = R + \frac{O}{K} \cdots\cdots\cdots\cdots\cdots\cdots \quad [2]$$

(Mehrens and Lehmann,1991, p300)

حيث إن:

R_C : تُشير إلى الدرجة الكلية المصححة للمفحوص على فقرات الاختبار.

R : تُشير إلى عدد الإجابات الصحيحة.

O : تُشير إلى عدد الفقرات المحذوفة.

K : تُشير إلى عدد البدائل التي تشتمل عليها الفقرة .

وتقوم هذه الطريقة لمعالجة التخمين على تشجيع المفحوص على ألاّ يُخمّن وذلك بمكافأته إذا لم يفعل ذلك، من خلال زيادة علامته الظاهرية بما يتناسب مع درجة تحفظه على الإجابة أو ابتعاده عن الإجابة بطريقة عشوائية عندما لا يُرجّح أي بديل من البدائل، وذلك بإضافة ذلك الجزء من العلامة الذي يتوقع أن يحصل عليه فيما لو خمّن إلى علامته الظاهرية حسب الصيغة

الرياضية السابقة، ويُلاحظ أن هذه الطريقة من التصحيح تزيد من علامة المفحوص الظاهرية من خلال مكافأته بنقاط إضـــافية على الفقرات التي قام بحذفها على افتراض أنه إذا كان قد حاول الإجابة عن الفقرة المحذوفة فإن احتمالية اختياره للإجابة الصـحيحة يسـاوي $\left(\frac{1}{K}\right)$، حيث K عدد البدائل، وتعتمد هذه الطريقة على افتراض أن جميع محاولات التخمين على الفقرات المحذوفة هي عشـوائية، وأن تأثير البعد النفسـي للحافز وهو المكافاة بإضـافة درجات مقابل الفقرات التي يقوم المفحوص بحذفها له تأثير قوي عليه بحيث يدفعه لعدم انتهاج سـلوك التخمين، أكثر من التهديد بالعقاب وحذف الدرجات(Kruz,1999).

بينما يرى أبو سـياف (Abu-Sayf,1977) أن هذه المعادلة فعليًا لا تقوم بالتصـحيح لأثر التخمين بل هي بالواقع تهدف إلى تشـجيع المفحوص على عدم التخمين من خلال مكافأته على الفقرات التي قام بحذفها لعدم معرفته بالإجابة الصحيحة.

وأشـار ايبل وفرزبي (Ebel and Frisibi,1991) إلى أنه عندما يتم تصـحيح درجات اختبارين من أثر التخمين بطريقتين مختلفتين، مثل معادلة التصـحيح لأثر التخمين(العقاب)، ومعادلة التصـحيح لأثر التخمين(المكأفاة)؛ فإن ناتج عملية التصـحيح يُعطي مجموعتين مختلفتين من الدرجات، ولكن مُعامل الارتباط بين هاتين المجموعتين من الدرجات يكون ارتباطًا تامًا.

ثالثا: طريقة معادلة التصحيح لأثر التخمين (المكافأة والعقاب)

تستند هذه الطريقة للتصحيح ضد أثر التخمين إلى الافتراضات الآتية:

أ. المُمتحن يمتلك الإجابة الصحيحة للفقرة.

ب. المُمتحن لا يعرف الإجابة الصحيحة، وبالتالي يحذف الفقرة.

ج. المُمتحن لا يعرف الإجابة الصـحيحة عن الفقرة، ويُجيب عنها بشـكل عشـوائي (Kruz,1999).

إذ يرى أصـحاب هذه الطريقة أن الطرق السـابقة لم تأخذ بعين الاعتبار جميع الافتراضـات السـابقة، ويُلاحظ أن الطريقة الأولى إعتمدت على الإجابات الصـحيحة والإجابات الخطأ، بينما اعتمدت الطريقة الثانية على الإجابات الصـحيحة والإجابات المحذوفة، لذلك جاءت هذه الطريقة لتجمع الافتراضـات الثلاثة السـابقة بمُعادلة واحدة تُعرف بمعادلة التصـحيح لأثر التخمين (المكافأة والعقاب)، وتمتاز هذه المعادلة عن سـابقتيها بأنها تكافئ المفحوص على الفقرات المتروكة بدون

إجابة مبنية على التخمين، وتعاقبه على الفقرات التي يُتوقع أنه قد خمّن الإجابة عنها، أي تجمع بين المكافأة والعقاب، وتنصّ على أن العلامة المصححة بهذه الطريقة هي:

$$\text{علامة المفحوص} = \text{عدد الاجابات الصحيحة} + \frac{\text{عدد الاجابات المحذوفة}}{\text{عدد البدائل}} - \frac{\text{عدد الاجابات الخطأ}}{\text{عدد البدائل} - 1}$$

(Chiu and Camili, 2012, p79)

تمتاز هذه المعادلة بأنها توازن بين المبالغة في درجة المفحوص الناتجة عن إضـــافة علامات له بدل الفقرات التي لم يُجب عنها لعدم لجوئه للتخمين، وانخفاض علامته لعدم مراعاة تصـــحيح التخمين الناتج عن المعرفة الجزئية أو التخمين المنطقي القائم على المعرفة الجزئية (Ebel and Frisibi,1991).

ويرى لســج وفالكي و ســابي (Lesage,Valcke and Sabbe,2013) أن المنطق وراء هذه المعادلة هو عدم تشــجيع المفحوص لانتهاج ســلوك التخمين عند أدائه على فقرات الاختبار الموضوعي/ الاختيار من متعدد، ولكنها لم تحل مشكلة التخمين نفسها؛ بل أضافت مشكلة جديدة إذ يختلف المفحوصــون في مدى ممارســتهم لسـلوك التخمين، فبعضــهم لديه الجرأة لانتهاج سـلوك المخاطرة والإقدام على التخمين أكثر من غيرهم من المفحوصـــين؛ وبالتالي فإن الطلبة الذين يلجأون للتخمين قد يحصـــلون على درجات أكثر من زملائهم من مســتوى القدرة نفســه الذين يلتزمون بالتعليمات ويخشون المخاطرة، مما قد يؤثر على صدق الاختبار وثباته.

وعموما تضـمن الأدب النظري الذي كتب حول عيوب اسـتخدام معادلات التصـحيح لأثر التخمين الأمور الآتية:

١- ليس من الضروري أن تكون إجابات المفحوص الخطأ دليلاً على التخمين، فقد تشير هذه الإجابات إلى عدم معرفته فعلاً للجواب الصحيح، واستخدام معادلة التصحيح في هذا الحال للقيام بخصم جزء من علامة المفحوص يؤثر عليه معنويًا ونفسيًا.

٢- إن زيادة جزء من العلامة للمفحوصين الذين لا يُخمّنون يُعطي بعضهم فرصة لكسب علامة غير مشروعة، وفي هذا تعارض مع مبدأ القياس الذي يؤكد ضرورة معرفة الفروق الفردية بين المفحوصين، ويُعدّ خللاً في قياس الوضع الحقيقي لهم وكذلك معرفة القدرة الحقيقية لكل منهم.

٣- عدم جدوى استخدام معادلة تصحيح أثر التخمين في حال إجبار جميع المفحوصين على إجابة جميع الأسئلة حيث أن رتب درجاتهم تبقى نفسها قبل تطبيق معادلة التصحيح وبعد تطبيقها (Fray,1988).

رابعاً: طريقة اختيار أكثر من بديل (مكافأة المعرفة الجزئية)

تم اقتراح هذه الطريقة من قِبل بوش (Bush,2001) إذ يسمح للمفحوص هنا باختيار إجابة واحدة فقط لكل فقرة، عندما يكون متأكدًا من الإجابة الصحيحة، وعندما لا يكون متأكدًا من إجابته فيمكن اختيار أكثر من بديل، وبناءً على ذلك فإن هذه الطريقة تفترض ما يلي:

١- يعرف المفحوص الإجابة الصحيحة ويتم اختيارها، ويحصل على الدرجة كاملة.

٢- المفحوص غير المتأكد من إجابته وحصر الإجابة الصحيحة بين بديلين تكون الدرجة التي يحصل عليها لهذه الفقرة مساوية لـ ((3-1)/3=67%)

٣- المفحوص غير المتأكد من إجابته، وحصر الإجابة الصحيحة بين ثلاثة بدائل، فإن درجته المستحقة تكون مساوية لـ ((3-2)/3= 33%) من علامة الفقرة.

ويرى بوش أن هذه الطريقة تمكن المفحوص من استغلال المعرفة الجزئية التي يمتلكها عند إجابته على فقرات الاختيار من متعدد، وتُحد من التخمين العشوائي.

خامساً: طريقة إجراء الحذف لكومبس وملهولند وومر.

(The Elimination Procedure of Coombs, Milholland, and Womer)

وضع هذه الطريقة كل من كومبس و ملهولند و وومر(Coombs , Milholland and ,1956 Womer) ، وتُشير تعليمات هذه الطريقة إلى أن المفحوص خلال إجابته عن فقرات اختبار الاختيار من متعدد يقوم باستبعاد البدائل غير الصحيحة، وبالتالي فهو يُكافأ بعلامة واحدة لكل بديل غير صحيح يستبعده، وتتم معاقبته بخصم (K-1) من العلامات؛ إذ تمثل (K) عدد البدائل للفقرات التي يقوم المفحوص باستبعادها(Bush,2001) .

وأهم ما يُميز هذه الطريقة أنها تفترض وجود مستويات مختلفة من المعرفة لدى المفحوصين، فالمفحوص الذي يمتلك معرفة تامة (Full knowledge) يقوم باستبعاد جميع البدائل الخطأ، بينما المفحوص الذي يمتلك معرفة جزئية (Partial knowledge) يقوم باستبعاد مجموعة جزئية من الإجابات الخطأ، أما المفحوص الذي يمتلك معلومات غير صحيحة جزئيًا (Partial Misinformation) فهو يستبعد الإجابة الصحيحة، ومجموعة جزئية من البدائل الخطأ،

وأخيرًا فإن المفحوص الذي يفتقر للمعلومات الصحيحة تمامًا يقوم باستبعاد الإجابة الصحيحة فقط (Ben-Simon, Budescu and Nevo, 1997).

سادسا: طريقة تصحيح المموهات(البدائل) Distraction –Scoring

اقترح هذه الطريقة دودين (Dodeen, 2005) وما يُميّز هذه المعادلة أنها تأخذ بعين الاعتبار مستوى جاذبية المموه(البديل) في حساب درجة السؤال النهائية، وتفترض أنّ التخمين ليس عشوائيًا دائمًا، وأن الاختيار الأكثر جاذبية هو الاختيار الأقرب إلى الصواب من ضمن قائمة البدائل في سؤال الاختيار من متعدد والصيغة الرياضية لهذه الطريقة هي:

$$F = R - \frac{\sum_{r=1}^{K} M_r}{A-1} \qquad \text{......................[3]}$$

حيث إن: (Dodeen, 2005, p14)

F: الدرجة النهائية المُصححة.

R: عدد الفقرات التي تمت إجابتها إجابة صحيحة.

A: عدد بدائل الفقرة.

r : رقم الفقرة .

K: عدد الفقرات المكونة للاختبار.

M_r: معامل رياضي وتتحدد قيمته بـ ($\frac{1}{A-1}$) إذا اختار المفحوص البديل الأكثر اختيارًا (جاذبية) من قبل المفحوصين الآخرين، وتساوي قيمته ($\frac{2}{A-1}$) إذا اختار البديل الثاني الأكثر جاذبية، وتساوي (١) إذا اختار المفحوص أقل البدائل (المموهات) اختيارًا.

وقد أجرى دودين دراسة من أجل مقارنة المعادلة الجديدة Distraction Scoring (DSF) Formula بالمعادلة التقليدية لتصحيح أثر التخمين الشائعة الاستعمال، كما قام بمقارنة فعالية المعادلتين؛ ولاسيما تأثير كلّ منهما على درجات الاختبار، وعلى الثبات لتلك الدرجات؛ من خلال تطبيق المعادلتين على نتائج اختبارين معياريين في جامعة الإمارات العربية المتحدة، وقد دلّت النتائج على أفضلية استخدام معادلة تصحيح المموهات في مكافأة الطلبة على حسن اختيارهم للبديل

الأقرب إلى الصواب أو الأكثر صحة عندما يكونون غير متأكدين تمامًا من الجواب الصحيح، بالإضافة إلى ذلك فقد أدى تطبيق هذه المعادلة إلى زيادة ثبات نتائج الاختبارين عندما حُسبت على عينات فرعية مشتقة من العينة الكلية لكل اختبار.

سابعاً: طريقة استخدام البديل "لا أعرف الإجابة "

تقوم هذه الطريقة على استخدام البديل "لا أعرف الإجابة " بين بدائل فقرات الاختبار مع استخدام طريقة معاقبة التخمين، وتستخدم هذه الطريقة بشكل كبير في الكليات الطبية، وهي تساهم بشكل فعال في الحد من أثر التخمين على درجات الاختبار الموضوعي/ الاختيار من متعدد، كما وتستخدم هذه الطريقة في كل من الدنمارك وكندا وألمانيا، ويرى كل من مجتجنز و مامرنبم و وهجنبوم و ايفرز (Muijtjens,Mamerenp, Hoogenboom and Evers,1999) أن الاختبارات الموجهة لقياس مدى تحسن الطلبة عبر المحتوى الدراسي كاملاً لا بدّ أن تتضمن مثل هذا البديل؛ لأنه من المتوقع عدم تمكن بعض الطلبة من الوقوف على كافة النتاجات التي يسعى الاختبار لتقييمها، أما ويد وهاريسون وهولاند وماتك وريكتس وواس (Wade, Harrison, Hollands, Mattick, Ricketts, Wass, 2011) فقد توصلوا من خلال دراسة تجريبية قاموا بها إلى أهمية البديل "لا أعرف الإجابة " في التحكم بأثر التخمين، كما وجدوا أن الطلبة الذين تقدموا لاختبار لا تتضمن فقراته البديل "لا أعرف الإجابة " يرون أن الحظ والتخمين ساهما بشكل كبير في نجاحهم في الاختبار أكثر من المعرفة التي يمتلكونها، وفي هذا السياق يرى بورتون (Burton,2002) أن استخدام البديل "لا أعرف الإجابة " يزود المفحوصين والمعلمين بالتغذية الراجعة المناسبة حول الاختبار وحول جوانب القصور لدى المفحوصين ليصار لاحقًا إلى معالجتها في المواقف التعليمية داخل غرفة الصف، بينما يرى كل من اسبونزا وجاردازبل (Espinoza and Gardeazabal,2010) أن إضافة البديل " لا أعرف الإجابة " إلى بدائل فقرات الاختبار يقود إلى خطأ قياس من خلال التحيز ضد المفحوصين الذين يميلون للمخاطرة والذين يختارون بديلاً آخر غير البديل "لا أعرف الإجابة " حتى عندما لا يمتلكون المعرفة التامة بالإجابة الصحيحة.

ومع نهاية استعراض الأدب النظري المتعلق بالإجراءات والطرق المستخدمة في إطار النظرية الكلاسيكية في القياس للتغلب على مشكلة التخمين في الاختبارات الموضوعية/الاختيار من متعدد، ومع اختلاف الآليات والطرق التي وضعت والتي تراوحت ما بين الحد من سلوك التخمين من خلال معاقبة المفحوص بخصم جزء من درجته الكلية، إلى مكافأة المفحوص لعدم لجوئه للتخمين، إلى طرق سعت للدمج بين هاتين الطريقتين (المكافأة والعقاب).

فقد سعت طرق اخرى إلى الاهتمام بالمعرفة الجزئية التي يمتلكها المفحوص؛ إذ نظرت للتخمين على أنه ليس دائما عشوائيًا، والقاسم المشترك بين جميع هذه الطرق التي ظهرت في إطار النظرية الكلاسيكية في القياس انها جميعها سعت للتغلب على مشكلة التخمين، والوقوف على المستوى الحقيقي للمفحوص، وتحسين الخصائص السيكومترية للاختبار من صدق وثبات، على الرغم منْ أن كل طريقة من هذه الطرق لها إيجابيات وسلبيات سواء على خصائص الاختبار أو على المفحوصين.

معالجة التخمين في نظرية الاستجابة للفقرة

الواقع أن المطّلع على الأدب التربوي الذي تناول قضية التخمين في إطار نظرية الاستجابة للفقرة يجد أن هناك عددًا من النماذج والإجراءات التي تم اتباعها للتعامل مع مشكلة التخمين، والتي طُوِّرت في إطار النموذج اللوجستي ثلاثي المعلمة.

وبشكل عام ينظر إلى التخمين في إطار نظرية الاستجابة للفقرة على أنه الظاهرة التي تسمح للمفحوص ذي القدرة المنخفضة أن يجيب إجابة صحيحة عن الفقرة التي صعوبتها أعلى من مستوى قدرته، ممّا يجعل ذلك يتناقض مع ما هو متعارف عليه من أن المفحوصين ذوي القدرة المرتفعة هم فقط من يجيبون إجابة صحيحة عن فقرة توازي قدرتهم أو تقل عنها (Hambleton and Swaminathan, 1985).

من هنا فإن العاملين في إطار نظرية الاستجابة للفقرة اقترحوا عددًا من النماذج والبدائل للتعامل مع مشكلة التخمين في الاختبارات الموضوعية، حيث جرى التعامل معه في إطار النموذج اللوجستي الثنائي المعلمة ونموذج راش عن طريق الافتراض بانعدامه، إذ يتم تحديد الفقرات التي تكون إجابتها لا تنسجم مع نمط استجابة المفحوص ويتم حذفها لعدم مطابقتها مع النموذج، كما يحدث في حالة إجابة المفحوص عن فقرات صعبة، ويخطئ في الإجابة عن الفقرات السهلة، ويُشير كل من واينر ورايت(Wainer and Wright,1980) إلى طريقة أخرى للتعامل مع التخمين؛ والتي تقوم على حذف تلك الفقرات التي تُظهر أن إجابة المفحوص عنها لم تكن منطقية بالنظر إلى نمط استجابته على الفقرات الأخرى، ويتم الاكتفاء بالفقرات المتبقية لتقدير قدرته.

أما في إطار النموذج اللوجستي الثلاثي المعلمة فقد تم التعامل مع مشكلة التخمين عن طريق تقديرها بمعلمة أطلق عليها معلمة التخمين (c_i) ولكن يبدو أن هناك وجهات نظر مختلفة حول طرق تقديرها والتعامل معها؛ فهناك من يفترض أن لكل المفحوصين حتى المتفاوتين في قدرتهم المعرفية نفس القدرة على التخمين بنجاح، في حين أن البعض الآخر يفترض أن التخمين يرتبط بمستوى القدرة

لدى المفحوص، ولعل عدم الاتساق هذا ربما يعود إلى الاختلاف في تحديد ما يُسمى بالتخمين المنطقي؛ الذي يعتقد الكثيرون أنه يحصل عندما لا يمتلك المفحوص المعرفة الكافية للإجابة عن فقرة الاختبار إجابة صحيحة، إذ يُعد التخمين المنطقي جزءًا من عملية التخمين الكلية التي يقوم بها الفرد، ممّا قاد إلى صعوبة الفصل بين التخمين المنطقي والتخمين العشوائي، حيث يرى هان (Han,2012) أن التخمين المنطقي المبني على المعرفة الجزئية هو جزء من عملية حل المشكلات التي يستخدمها المفحوص خلال إجابته عن فقرات الاسئلة الموضوعية /الاختيار من متعدد.

واقترح هان أن كل من معلمة التمييز ومعلمة التخمين للفقرة ومعلمة القدرة للمفحوصين في النموذج اللوجستي الثلاثي المعلمة لا تُعتبر معالم ذات دقة كافية، ويرى أن يتم تثبيت معلمة التخمين في النموذج اللوجستي الثلاثي المعلمة لتصبح مساوية لـ ($\frac{1}{\text{عدد بدائل الفقرة}}$)، ممّا يُعطي مطابقة أفضل للبيانات.

ولمحاولة معالجة التخمين في إطار الأسـئلة الموضـوعية/الاختيار من متعدد قام عدد من الباحثين بوضـع واقتراح عدد من النماذج الرياضـية، والآليات بهدف التخلص أو التقليل من أثر التخمين على فقرات اختبار الاختيار من متعدد، ففي هذا السـياق عمل والر (Waller,1989) على تقديم نموذج لإزالة أثر التخمين العشـوائي؛ وهو ما يُعرف بنموذج (ARRG) Ability Removing Random Guessing إذ يتم التعامل مع مشـكلة التخمين من خلال التركيز على التفاعل بين الفقرة والمفحوص، فيفترض هذا النموذج أن المفحوص الذي يُخمّن عشـوائيًا خلال إجابته عن فقرات الاختبار يقوم بذلك فقط على الفقرات التي يعتبرها صـعبة جداً بالنسـبة له، حيث يتم حذف هذه الفقرات عند تقدير معلمة القدرة (θ) للمفحوص، وبهذا لن يتأثر تقدير معلمة القدرة سـواء لجأ المفحوص للتخمين أم لم يلجأ، وتتمثل الصيغة الرياضية لهذا النموذج في المعادلة الآتية:

$$P_{ij} = \frac{1}{1+\exp\left[a_j(b_j-\theta_i)\right]} \qquad P_{ij} > P_c, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots[4]$$

(Waller,1989, p.234)

حيث إن:

P_c: نقطة القطع والتي هي أقل أو تساوي (1/A)، و(A) تمثل عدد بدائل الفقرة.

P_{ij} : احتمالية إجابة المفحوص (i) على الفقرة (J) إجابة صحيحة.

a_j : معلمة تمييز الفقرة (j).

b_j : معلمة صعوبة الفقرة (j).

θ_i : معلمة القدرة للمفحوص (i).

بشكل عام يعمل هذا النموذج وفق الاجراء الآتي:

$$P_{ij} = \begin{cases} \dfrac{1}{1+\exp[a_j(b_j-\theta_i)]} & \text{for } P_{ij} > P_c \\ 0 & \text{for } P_{ij} \leqslant P_c \end{cases}$$

(Waller,1989, p.234)

إذ يتم تقســـيم الفقرات لكل مفحوص إلى قســـمين: تلك التي احتمالية الإجابة عنها بدون تخمين أعلى من احتمالية الإجابة عنها بالتخمين، وتلك التي احتمالية الإجابة عنها أقل من أو تســاوي احتمالية الإجابة بالتخمين (P_c) ، ويســتخدم النموذج فقرات المجموعة الأولى فقط لتقدير قدرة المفحوص.

بالرغم من أن النموذج الســابق اســتخدم عددًا أقل من الفقرات لتقدير قدرة المفحوص مقارنة مع النموذج اللوجســتي الثنائي المعلمة، إلا أن الدقة في التقدير كانت أعلى، ولكنه مقارنة مع النموذج اللوجستي الثلاثي المعلمة يُعطى مطابقة أقل للبيانات(Gao,2011) .

نماذج نظرية استجابة الفقرة المعدلة Modified IRT Models

يفترض كلّ من النموذج الأحادي المعلمة والنموذج اللوجســتي الثنائي المعلمة في نظرية الاســتجابة للفقرة أن معلمة التخمين تســاوي صــفرًا، في حين أنه على أرض الواقع في الاختبارات الموضوعية/ الاختيار من متعدد يصعب الحصول على معامل تخمين مساوٍ للصفر، وبالتالي جاءت النماذج المعدلة لنظرية الاســتجابة للفقرة التي افترضــت وجود قيمة تخمين ثابتة غير مســاوية للصــفر تُضــاف للصــيغة الرياضــية للنموذج، وأن احتمالية إجابة الفقرة إجابة صــحيحة تســاوي ($\frac{1}{\text{عدد البدائل}}$) ، واحياناً قد يمتلك المفحوص معرفة جزئية حول ما تقيسه الفقرة مما يجعله قادرًا على حذف بعض البدائل وبالتالي ترتفع احتمالية الإجابة بشكل صحيح عن القيمة ($\frac{1}{\text{عدد البدائل}}$). وقد اقترح وارم (Warm,1978) أن قيمة معلمة التخمين تقريبًا تساوي (0.05 - ($\frac{1}{A}$)) في حالة وجود أربعة بدائل للفقرة حيث (A) تمثل عدد البدائل لتصــبح (٠,٢٥ - ٠,٠٥) أي (٠,٢٠) وفي حالة

وجود خمسة بدائل للفقرة تصبح معلمة التخمين مساوية (٠,١٥). ومما يُميز النماذج المعدلة عدم الحاجة إلى عينات كبيرة من المفحوصين كما في النماذج اللوجستية التقليدية؛ ممّا يجعل التطبيق العملي لها أكثر سهولة(Parshall, Kormrey and Chason, 1997).

وعموما تقسم النماذج المعدلة لنماذج نظرية الاستجابة للفقرة إلى قسمين رئيسين، هما:

١- **النموذج الأحادي المعلمة المعدّل:**

جرى تعديل النموذج اللوجستي أحادي المعلمة بتحويله إلى ما يُعادل النموذج اللوجستي ثنائي المعلمة؛ وذلك بافتراض أن معاملات تمييز الفقرات تأخذ قيمة واحدة يتم تقديرها، بإضافة ثابت غير مساوٍ للصفر بدلا من معلمة التخمين، ويكون حجم العينة المطلوب لتقدير معالم الفقرات وقدرات المفحوصين مماثلاً لحجم العينة المطلوب في النموذج الأحادي المعلمة التقليدي.

وتتمثل الصيغة الرياضية للنموذج أحادي المعلمة المعدّل بالشكل الآتي:

$$p_i(\theta) = C + (1 - C)\frac{e^{D(\theta - bi)}}{1 + e^{D(\theta - bi)}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots[5]$$

(Barnes and Wise,1991 ; Baldwin, 2006)

حيث إن:

P(θ): تمثل احتمال إجابة المفحوص ذي القدرة (θ) (الذي اخُتير عشوائيًا)على الفقرة (i) إجابة صحيحة.

D: قيمة ثابتة تساوي (١,٧٧).

c_i: معلمة تخمين الفقرة، ثابتًا غير صفري.

bi: معلم الصعوبة للفقرة (i) .

e: الأساس اللوغاريتمي الطبيعي ويساوي (٢,٧١٨).

٢- **النموذج اللوجستي الثنائي المعلمة المعدَّل:**

يتمثل التعديل الذي أدخل على هذا النموذج بافتراض أن قيمة معلمة التخمين ثابتة وغير مساوية للصفر، وتساوي في حدّها الأعلى ($\frac{1}{\text{عدد البدائل}}$)، ومن هنا ينتج عن استخدام هذا النموذج مطابقة أفضل للبيانات، ودقة في تقدير كل من معالم الفقرات ومعلمة القدرة، وأما حجم العينة المطلوب من المفحوصين لإعطاء تقديرات دقيقة فهو مماثل لحجم العينة المطلوب في النموذج

الثنائي المعلمة التقليدي، الأمر الذي يجعل التطبيق العملي لهذه النماذج أقل صعوبة، خاصة عند وجود أعداد قليلة من المفحوصين.

وتكون المعادلة الرياضية للنموذج ثنائي المعلمة المعدّل هي:

$$p_i(\theta) = C + (1 - C)\frac{e^{Dai(\theta - bi)}}{1 + e^{Dai(\theta - bi)}} \qquad ....[6]$$

.(Baldwin 2006 ;Barnes and Wise ,1991; Sireci,1992)

حيث إن:

P(θ): تمثل احتمال إجابة المفحوص ذي القدرة (θ) الذي اختير عشوائيًا على الفقرة (i) إجابة صحيحة.

D: قيمة ثابتة تساوي (١,٧٧)

a_i: معلمة تمييز الفقرة.

C: معلمة تخمين الفقرة، وتساوي ثابت غير صفري.

b_i: معلمة الصعوبة للفقرة (i).

e: الأساس اللوغاريتمي الطبيعي ويساوي (٢,٧١٨).

θ: : معلمة القدرة للمفحوص.

لقد جرت دراسات كثيرة تناولت النماذج المعدّلة منها: دراسة البنزي وفورسيث (Albanese and Foresyth, 1984) التي قارنا فيها بين النموذج الأحادي المعلمة والنموذج اللوجستي الثنائي المعلمة، والنموذج اللوجستي الثنائي المعلمة المُعدّل، حيث كانت قيمة التخمين تساوي ثابتًا غير صفري، وأظهرت نتائج الدراسة أن النموذج اللوجستي الثنائي المعلمة المعدّل أفضل مطابقة للبيانات من النماذج الأخرى(النموذج اللوجستي الأحادي المعلمة والنموذج اللوجستي الثنائي المعلمة) المستخدمة في الدراسة، كما كانت دقة التقديرات لمعالم الفقرات وقدرات الأفراد أعلى عند استخدام النموذج اللوجستي الثنائي المعدّل.

النماذج المختلطة Mixture Model

تستخدم نماذج نظرية الاستجابة للفقرة التقليدية نموذجًا لوجستيًا واحدًا لتقدير معالم الفقرات ومعلمة القدرة للمفحوصين، أما النماذج المختلطة والتي وضعها يامتو (Yamamoto,1995)، فتحتوي صيغتها الرياضية على نموذج لوجستي بالإضافة إلى معاملات رياضية لتحديد فيما إذا لجأ المفحوص للتخمين أم لم يلجأ عند إجابته عن فقرات الأسئلة الموضوعية/الاختيار من متعدد.

والمعادلة التالية تُبين مثالاً على نموذج مختلط يدمج بين النموذج اللوجستي الأحادي المعلمة، واحتمالية استخدام المفحوص للتخمين والممثلة بالمقطع (c_i).

كما ؛

$$P(x_i=1|\ \theta, B_i\ , K)= (1 + \exp(-Da_i\,(\theta-b_i)))^m\ c_i^{m+1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots[7]$$

(Yamamoto, 1995, p.5)

K: آخر فقرة أجاب عنها المفحوص.

x_i : احتمال إجابة المفحوص إجابة صحيحة عن الفقرة (i).

m: معامل رياضي تتحدد قيمته بـ (-١) عندما $K \geq i$ ، وتساوي (صفر) عندما $i > k$

B_i: الاستجابة الثنائية وتأخذ القيمة (صفر) في حالة الإجابة الخطأ ، والقيمة (١) في حالة الإجابة الصحيحة.

D: قيمة ثابتة تساوي (١,٧).

e: الأساس اللوغاريتمي الطبيعي ويساوي (٢,٧١٨).

θ: معلمة القدرة للمفحوص.

b_i: معلمة الصعوبة للفقرة (i).

c_i : احتمالية الحصول على إجابة صحيحة من خلال التخمين.

حينما تكون قيمة (m= -1) فإن النموذج أعلاه يتحول إلى النموذج اللوجستي الثنائي المعلمة، وعندما تكون قيمة (m=0)، فهذا يعني أن المفحوص قد استخدم التخمين وبالتالي فإن النموذج يتحول من نموذج لوجستي ثنائي المعلمة إلى نموذج مختلط يدمج بين النموذج الثنائي المعلمة، والمعلمة (c_i) التي تمثل استخدام المفحوص للتخمين.

تُستخدم هذه النماذج عادة للتعامل مع مجموعات مختلفة من المفحوصين وفي هذه النماذج يمكن أن يكون هناك مجموعتان إحداهما تقوم بالتخمين والأخرى لا تقوم به، أو إحداهما مرتفعة الدافعية والأخرى منخفضة الدافعية، فالنمط الأول يمكن استخدام نماذج نظرية الاستجابة للفقرة معه، أما النمط الثاني فيحتاج نماذج مختلفة، وقد وجد اكسي (Xie,2002) أن النموذج المختلط يُعطي مطابقة أفضل للبيانات من نموذج راش لوحده، كما يعطي دقة تقديرات لكل من معلمة قدرة الفرد ومعالم الفقرة أفضل مما يُعطي النموذج اللوجستي الثلاثي المعلمة.

وفي هذا السياق أيضا قامت ستوك وكاو (Cao and Stokes,2008) بتطوير نموذجين يستندان إلى نماذج نظرية بييز للتعامل مع عدة أنماط أو سلوكات للتخمين هما: نموذج عتبة التخمين، ونموذج التخمين القائم على الصعوبة؛ وقد تم في هذه الدراسة استخدام كل من هذين النموذجين بالإضافة إلى كل من النموذج اللوجستي الثلاثي المعلمة والنموذج اللوجستي الثنائي المعلمة المعدّل إلى ثلاثي وفيما يلي استعراضٌ لهما:

١- **نموذج عتبة التخمين:** The IRT Threshold guessing Model

يفترض نموذج عتبة التخمين أن بعض المفحوصين يُمكن أن يجيبوا عن فقرات الاختبار على أساس معرفتهم إلى أن يصلوا إلى حدّ معين، فيقومون بالتخمين العشوائي بعد ذلك، ويقترح هذا النموذج تحديد عتبة التخمين للمفحوص لتحديد أي الفقرات التي ستقدم له، وبناءً عليها يتم تقدير معلمة القدرة له، كما يتضمن معلمة يُطلق عليها موقع الفقرة والتي تُحدد العتبة (الحد الفاصل) التي يبدأ عندها المفحوص باللجوء للتخمين للإجابة عن فقرات الاختبار.

وتتمثل الصيغة الرياضية لهذا النموذج، والتي تبين احتمالية إجابة المفحوص (i) على الفقرة (j) إجابة صحيحة بالآتي:

$$P(x_{ij}=1|\,\theta_i,\delta_j,\gamma_j,c_j,\alpha_i)=\frac{\exp[\gamma_j(\theta_i-\delta_j)-I(j>\alpha_i)(\gamma_j(\theta_i-\delta_j)-c_j)]}{1+\exp[\gamma_j(\theta_i-\delta_j)-I(j>\alpha_i)(\gamma_j(\theta_i-\delta_j)-c_j)]} \quad ....[8]$$

(Cao and Stokes,2008, p.5)

حيث إن:

(α_i): تشير إلى موقع معلمة الصعوبة للفقرة لكل مفحوص Examinee's Item Location (Threshold) Parameter.

(c_j): معلمة التخمين للفقرة.

γ_j : معلمة التمييز للفقرة.

θ_i : معلمة القدرة للمفحوص.

δ_j : معلمة صعوبة الفقرة.

أما الاقتران $I(j > \alpha_i)$ فيقارن بين احتمالية الإجابة على الفقرة (j) والحد الفاصـــل (α_i)، ويُعطي هذا الاقتران القيمة (١) في حالة كانت صـــعوبة الفقرة أكبر من العتبة مما يعني حصـــول عملية التخمين، ويُعطي القيمة (صــفر) في حال كانت قيمة العتبة أكبر من موقع الفقرة (j) أو تســاويها، وبهذا يصـــبح الجزء التالي من النموذج مســـاويًا للصـــفر، ويتحول النموذج المختلط إلى النموذج اللوجستي الثنائي المعلمة، وفيما يلي الصيغة الرياضية لهذا الاقتران:

$$I(j > \alpha_i) \;=\; \begin{cases} 1, & j > \alpha_i \\ 0, & j \le \alpha_i. \end{cases}$$

٢- **نموذج التخمين القائم على الصعوبة:** The IRT Difficulty-Based Guessing Model

يفترض هذا النموذج أن كل المفحوصـين سـواء الذين يسـتخدمون سـلوك التخمين أو الذين لا يسـتخدمونه سـيجيبون عن فقرات الاختبار، ولكن الطلبة الذين يسـتخدمون سـلوك التخمين يجيبون عن فقرات الاختبار السـهلة، ثم يسـتخدمون التخمين للإجابة عن بقية فقرات الاختبار، وبحسـب الصـيغة الرياضـية لهذا النموذج، فإن عتبة الصـعوبة أو فاصـل الصـعوبة ترتبط مع معلمة القدرة للمفحوص.

إذ تتمثل الصـــيغة الرياضـــية لنموذج التخمين القائم على الصـــعوبة التي تبين احتمالية إجابة المفحوص (i) على الفقرة (j) إجابة صحيحة تعطى وفق العلاقة الآتية:

$$p(x_{ij}=1|\theta_i,a_j,b_j,c_j,\delta_i) = \frac{\exp[a_j(\theta_i-b_j)-\eta_i I(b_j-(\theta_i+\delta_i))\,(a_j(\theta_i-b_j)-c_j)]}{1+\exp[a_j(\theta_i-b_j)-\eta_i I(b_j-(\theta_i+\delta_i))\,(a_j(\theta_i-b_j)-c_j)]} \quad \ldots\ldots\ldots[9]$$

(Cao and Stokes, 2008 , p9)

حيث إن:

a_j: معلمة تمييز الفقرة (j)

C_j: معلمة تخمين الفقرة (j)

b_j: معلمة الصعوبة للفقرة (j) .

e: الأساس اللوغاريتمي الطبيعي ويساوي (٢,٧١٨).

θ: معلمة القدرة للمفحوص (i)

η_i : معامل رياضي يساوي (١) إذا لجأ المفحوص للتخمين وتساوي (صفر) إذا لم يلجأ المفحوص للتخمين.

δ_i: معلمة تقيس عتبة الصعوبة التي يُخمّن عندها المفحوص .

والاقتران $I(bj- (\theta+ \delta_i))$ تكون قيمته مساوية (واحد) إذا كانت معلمة الصعوبة للفقرة (j) أكبر من قدرة المفحوص (i)، وتساوي (صفر) إذا كانت أقل وبذلك فإن قيمة المعلمة (δ_i) ستختلف من مفحوص لآخر.

٣- **نموذج جهد وقت الاستجابة:** Response –Time Effort Model (RTE)

اقترح وايز وديمارس (Wise and Demars, 2006) نموذجًا للتغلب على مشكلة التخمين في الاختبارات الموضوعية/الاختيار من متعدد، ويُسمى بنموذج جهد وقت الاستجابة إذ يتم حساب الزمن المبذول للإجابة عن كل فقرة في الاختبار وعن فقرات الاختبار ككل، ويُرمز لعملية الإجابة عن الاختبار بسلوك الحل Solution Behavior (SB)، ويُشار إلى سرعة التخمين بسلوك التخمين السريع ويُرمز لها بالرمز Rapid Guessing Behavior (RGB) وهو مؤشر إلى أن الجهد المبذول بالإجابة غير كافٍ، وعادة ما يُعبّر عن هذه العلاقة بالصيغة الرياضية الآتية:

$$RTE_j = \frac{\sum SB_{ij}}{k} \qquad \text{......................[10]}$$

(Wise and Kong, 2006,p4)

إذ تمثل (RTE_{ij}) مجموع الزمن الذي استغرقه المفحوص (i) للاستجابة على كل فقرة من فقرات الاختبار، في حين تمثل (K) عدد البدائل للفقرة، ويستخدم عادة هذا النموذج في الاختبارات المحوسبة؛ إذ يتم التمييز بين كل من الزمن المستغرق في الحل والزمن المستغرق في التخمين من خلال قيمة تمثل الحد الفاصل بينهما.

وقام كل من وايز وديمارس(Wise and Demars ,2006) بالتطوير على النموذج الســـابق والبناء على فكرة حساب زمن الاستجابة، وقدّموا نموذجًا جديدًا يُسمّى: نموذج الجهد المعدّل لزمن الاستجابة عن الفقرة Effort-Moderated Item Response Model (EMIRM) .

وقد تم التعبير عنه بالصيغة الرياضية الآتية:

$$P_i(\theta)=\left(SB_{ij}\right)\left(c_i+\left(1-c_i\right)\left(\frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}}\right)\right)+\left(1-SB_{ij}\right)\left(g_i\right) \quad ....[11]$$

(Wise and Demars , 2006,p24)

حيث إن:

$P_i(\theta)$: احتمال إجابة المفحوص إجابة صحيحة.

SB_{ij}: تعبر عن سلوك الحل (Solution Behavior) ويمثل الزمن المستغرق من قبل المفحوص (i) عند الاستجابة على الفقرة (j) بحيث يأخذ القيمة (1) إذا زاد زمن الاستجابة عن العتبة المحددة زمنًا مقدرًا للإجابة عن الفقرة، وبهذا يكون المفحوص قد سلك سلوك الحل، وإذا انتهى المفحوص من الإجابة عن الفقرة بزمن أقل من الزمن المحدد(العتبة) يكون المفحوص قد لجأ إلى التخمين في الإجابة عن الفقرة، وعندما تكون قيمة سلوك الحل (SB_{ij}) مساوية للواحد فإن النموذج أعلاه يتحول للنموذج اللوجستي الثلاثي المعلمة، وحينما تكون قيمة (SBij) مساوية للصفر فإن احتمالية الحصول على الإجابة الصحيحة هي دالة لـ (g_i) والتي ُتمثّل عدد البدائل للفقرة.

D : ثابت قيمته تساوي (١,٧٧).

b_i : معلمة صعوبة الفقرة.

θ: معلمة القدرة للمفحوص.

g_i : عدد بدائل الفقرة.

a_i : تمييز الفقرة.

وفيما يلي ملخص للنماذج اللوجستية التي وضعت لمعالجة التخمين في إطار نظرية استجابة الفقرة.

جدول رقم(١)

عرض للنماذج والإجراءات التي وضعت لمعالجة التخمين في إطار نظرية الاستجابة للفقرة.

الرقم	النموذج	واضعه	النظر للتخمين	معالجة التخمين
١	نموذج القدرة لإزالة التخمين العشوائي. ARRG ability removing random guessing.	Waller(1973)	التخمين عشوائي فقط ويحدث على الفقرات التي صعوبتها أعلى من قدرة المفحوص.	من خلال حذف تلك الفقرات التي يظهر من نمط استجابة المفحوص أنه لجأ للتخمين العشوائي.
٢	النموذج اللوجستي الأحادي المعلمة المعدّل:	Laura and Steven(1991)	تفترض هذه النماذج أنه لا يمكن القبول بمعلمة تخمين مساوية للصفر في نماذج نظرية الاستجابة للفقرة التقليدية.	يضيف هذا النموذج قيمة تتراوح بين (١/عدد البدائل) الى (ا/عدد البدائل – ٠,٢٠) للنموذج اللوجستي احادي المعلمة.
٣	النموذج اللوجستي الثنائي المعلمة المعدّل.	Laura and Steven(1991)	تفترض هذه النماذج انه لا يمكن القبول بمعلمة تخمين مساوية للصفر في نماذج نظرية الاستجابة للفقرة التقليدية.	يضيف هذا النموذج قيمة تتراوح بين (١/عدد البدائل) الى (ا/عدد البدائل – ٠,٠٥) للنموذج اللوجستي الثنائي المعلمة.
٤	نموذج تثبيت معلمة التخمين في النموذج اللوجستي الثلاثي المعلمة. Fixing the C Parameter in the Three-Parameter Logistic Model	Han(2012)	ينظر هذا النموذج للتخمين على أنه قيمة ثابتة عبر الاختبار لكل المفحوصين.	تثبيت معلمة التخمين بقيمة ثابتة تساوي (١/ عدد البدائل).

الرقم	النموذج	واضعه	النظر للتخمين	معالجة التخمين
النماذج المختلطة				
١	نموذج راش المختلط: Mixture-Rasch Model	Yamamoto(1995)	يراعي هذا النموذج أن مجموعة المفحوصين تختلف في مدى انتهاجها لسلوك التخمين، فهناك مجموعة لا تخمّن ومجموعة تلجأ للتخمين.	من خلال دمج نموذجين في نموذج واحد وإضافة معالم واقترانات تتحقق عبر معادلات رياضية من استخدام المفحوص للتخمين من عدمه، فإذا استخدم التخمين يعمل النموذج المختلط كاملاً، وإذا لم يُخمّن يعمل النموذج الثنائي فقط.
٢	نموذج عتبة التخمين: Threshold-Guessing Model.	Cao and Stokes(2008)	إن بعض المفحوصين يجيبون عن فقرات الاختبار على أساس معرفتهم إلى حد معين ثم يقومون بالتخمين على باقي الفقرات.	من خلال تحديد عتبة للتخمين، فالنموذج الرياضي يراعي الدمج بين النموذج الثنائي المعلمة وعتبة الفقرة.
٣	نموذج التخمين القائم على الصعوبة: Difficulty-Based Guessing Model	Cao and Stokes(2008)	المفحوص يجيب عن الفقرات بناءً على ما يمتلكه من معرفة ثم يخمّن عشوائيًا لاحقاً، بغض النظر عن مستوى الصعوبة لهذه الفقرات.	يضيف هذا النموذج معالم إضافية لتحري فيما إذا كان المفحوص مخمنا أم لا من نمط إجابته، فإذا استخدم المفحوص التخمين يتم تقدير قدرته باستخدام النموذج المختلط، وإذا لم يخمّن يُفعل النموذج اللوجستي الثنائي المعلمة.
٤	نموذج جهد وقت الاستجابة، Response-Time Effort Model	Wise(2006)	يميز هذا النموذج بين سلوك الحل وبين التخمين؛ فاذا قلّ الزمن الذي يستغرقه المفحوص للإجابة عن فقرة ما عن قيمة معينة فإن المفحوص يكون استخدم التخمين السريع(العشوائي).	الصيغة الرياضية لهذا النموذج تدمج بين النموذج الثلاثي المعلمة واقتران سلوك الحلّ، يُستخدم في الاختبارات المحوسبة، ويعتمد على تقدير زمن الاجابة للفقرة من المفحوص ومقارنته مع قيمة معينة لمعرفة فيما إذا استخدم المفحوص التخمين أم لا.

تقدير معلمة القدرة للمفحوصين ومعالم الفقرات.

يتم تقدير قيمة معلمة القدرة ومعالم الفقرات في إطار نظرية الاستجابة للفقرة باستخدام عددًا من الطرق والنماذج التي تم تطويرها، ولكن بشكل عام تمثل طريقة الأرجحية العظمى والطريقة البيزية أبرز الطرق التي طورّت في سياق نظرية الاستجابة للفقرة لتقدير معالم الأفراد والفقرات.

أولاً: طريقة الأرجحّية العظمى (MLE) Maximum Likelihood Estimation:

تعُد هذه الطريقة من أكثر طرق التقدير شهرة وشيوعًا، وخاصة في الاختبارات التكيفية المحوسبة وفي بنوك الفقرات وغيرها من تطبيقات نظرية الاستجابة للفقرة، حيث يتم تقدير معلمة القدرة من خلال نمط استجابة المفحوص على مجموعة من فقرات الاختبار، وتأخذ هذه الإجابة القيم (١) إذا كانت صحيحة، و(صفر) إذا كانت خطأ، حيث يُرمز لنمط إجابة المفحوص بالرمز (U)، وتحسب دالة الأرجحية العظمى من خلال إيجاد مجموع حاصل ضرب نمط إجابة المفحوص عن فقرات الاختبار كما في المعادلة الآتية(Hambelton,1989).

$$L\left(\mathbf{U}_j \mid \theta_j\right)=\prod_{i=1}^{n} P\left(u_{ij} \mid \theta_j\right)^{u_{ij}} Q\left(u_{ij} \mid \theta_j\right)^{1-u_{ij}} \qquad \text{......................[12]}$$

(Reckase,2009, P.140)

حيث إن:

U_{ij} : إجابة المفحوص (j) من مستوى القدرة (θ) على الفقرة (i) حيث تأخذ القيمة (١) إن كانت الإجابة صحيحة، والقيمة (صفر) إن كانت خطأ.

L(Uj| θ) :هي دالة الأرجحّية لإجابة المفحوص (j) من مستوى القدرة (θ) عن الفقرة (i)

P(Uij|θ): احتمال إجابة المفحوص (j) من مستوى القدرة (θ) عن الفقرة (i) إجابة صحيحة.

Q(Uij|θ): احتمال إجابة المفحوص (j) من مستوى القدرة (θ) عن الفقرة (i) إجابة خطأ

n : عدد فقرات الاختبار.

وتبدأ هذه الطريقة كما اشـــار لورد (Lord,1980) بافتراض قيمة أولية لقدرة المفحوص (θ_s)، وتُحســب دالة الأرجحّية عند أكثر من قيمة لمعلمة القدرة (θ) ، وقيمة معلمة القدرة (θ) التي تعطي أكبر قيمة لدالة الأرجحّية العظمى هي قدرة المفحوص، ويتطلب ذلك إيجاد المشــتقة الأولى لدالة الأرجحية بالنسبة لكل معلم من معالم النموذج المعين، ثم مساواة هذه المشتقات بالصفر وحلّ نظام المعدلات غير الخطيّة الناتجة باستخدام أساليب التحليل العددي مثل أسلوب نيوتن رافسون، إذ يقوم هذا الأســلوب بالبحث بطريقة متكررة؛ فيقوم بالتتابع بتحســين القيم التقديرية لهذه المعالم بحيث تحقق القيمة النهائية محك التقارب، ولتقدير معلمة القدرة للمفحوص تُستخدم المعادلة الآتية:

$$\theta_{S+1} = \theta_S + \frac{\sum_{i=1}^{n} - a_i [u_i - p_i(\theta_S)]}{\sum_{i=1}^{n} a_i^2 p_i(\theta_S) Q_i(\theta_S)} \quad \ldots\ldots\ldots\ldots\ldots\ldots [13]$$

حيث إن:

θ_{S+1}: معلمة القدرة النهائية.

(θ_s): معلمة القدرة الأولية للمفحوص.

u_i: وتساوي (صفر) للإجابة الخطأ و (واحد) للإجابة الصحيحة.

$Pi(\theta_s)$: احتمال إجابة المفحوص الذي يُختار عشـوائيًا من مسـتوى قدرة (θ_s) الفقرة (i) إجابة صحيحة .

$Qi(\theta)$: وتُشير إلى احتمالية أن يجيب المفحوص الذي اختير عشوائيًا من مستوى القدرة (θ_s) الفقرة خطأ.

(Baker, 2004, P.87)

كما يتم حساب $P_i(\theta)$ لكل فقرة من فقرات الاختبار الأولية عند مستوى القدرة (θ =1) وهي القدرة الأولية المفترضـة للمفحوص ليتم تعويضـها في المعادلة، ويتم تكرار هذه الحسابات أكثر من مرة للحصــول على قيمة قريبة من الصــفر للحدّ الأيمن من المعادلة وبذلك تقترب قيمة ($\theta s+1$) من (θs)، وعندها تتوقف عملية التقدير.

تتصــف طريقة الأرجحية العظمى لتقدير معلمة القدرة للأفراد ومعالم الفقرات (الصــعوبة والتمييز والتخمين) بالقوة والاتســاق، بمعنى أنه كلما زاد عدد الفقرات والمفحوصــين اقتربت (θ) المقدرة من (θ) الحقيقية، كما ويكون توزيع (θ) المقدرة قريبًا من التوزيع الســويّ، كما لا تحتاج هذه الطريقة إلا لنمط إجابة المفحوص كي يتم تقدير قدرته وبالتالي لا تحتاج إلى معلومات مســبقة

عن توزيع معلمة القدرة في مجتمع المفحوصين، ومن جهة أخرى تتمثل محددات هذه الطريقة أنه عندما يجيب المفحوص عن جميع الفقرات إجابة صحيحة فإن قيمة (θ) المقدرة تساوي (+∞) وعندما يجيب إجابة خطأ عن جميع الفقرات فإن قيمة (θ) المقدّرة تساوي (-∞) وبالتالي يصعب الحصول على قيمة قصوى للأرجحية في هذه الحالة، ويتم معالجة هذه المشكلة في الاختبارات المبنية على القياس التكيفي المحوسب (CAT) من خلال وضع تقديرات عليا ودنيا لمثل هذه الأنماط من الإجابة مثل (4+ ,4-) على التوالي. كما لا ينتج في حالة النماذج اللوجستية الثلاثي المعلمة والثنائي المعلمة تقديرات ثابتة للأرجحية القصوى لمعالم الفقرات ولمعلمة القدرة للمفحوصين (Thompson, 2009).

وهناك طرق بديلة لتقدير كل من معالم الفقرات وقدرات المفحوصين قام بتطوير إجراءتها عدد كبير من العلماء وقدّمت حلاً للتقديرات التي لم تستطع طريقة الارجحية القصوى التعامل معها، وفيما يأتي عرضًا لهذه الطرق.

ثانيا الطرق البيزية (Bayesian Methods):

تعود طريقة بييز (Bayes) في الاستدلال الإحصائي للعالم البريطاني توماس بييز(Bayes Thomas) عام ١٧٦٣م وقد اسُتخدمت أساسًا في التعامل مع الحوادث الاحتمالية المنفصلة (Reckase,2009)، وتُعّد من الطرق التي تستخدم الأساليب الإحصائية الشائعة لتقدير معالم الفقرات ومعلمة القدرة للمفحوصين في نماذج نظرية الاستجابة للفقرة، إذ تستند هذه الطريقة إلى اعتبار أن أي معلمة يتم تقديرها هي متغير عشوائي، ولها توزيع احتماليّ مُعين، يُسمى التوزيع القبلي (Prior Distribution)، وذلك من خلال الأخذ بعين الاعتبار المعلومات السابقة للمعلمة غير المعروفة في عملية التقدير، وهذه المعلومات غالبًا ما يتم تحديدها من الباحث مستندًا إلى خبرة سابقة أو إلى تقديره الشخصي، وفي الغالب تفترض هذه الطريقة أن التوزيع سوي، أو تفترضه ملتويًا نحو اليمين أو نحو اليسار(Bellhouse,2004; Fox,2010) .

وفي هذا السياق يرى لورد (Lord,1984) أن التقدير البييزي يزود بتقديرات لمعالم الفقرات ومعلمة القدرة للمفحوصين بشكل أدق من طريقة الأرجحية العظمى والسبب بحسب لورد أن التقدير البييزي يستند إلى معلومات أكثر، وحتى في غياب اية معلومات عن شكل التوزيع القبلي للبيانات فإنه يُمكن للطرق البييزية استخدام البيانات الموجودة لتوليد معلومات منها، ويضيف لورد أن طريقة التقدير البيزية تستخدم الوسط البعدي للبيانات كمقدر للمعالم، ممّا يقلّل من قيمة الجذر التربيعي لمتوسط الخطأ، وتُستخدم هذه الطريقة في العادة عندما لا يكون بالإمكان تطبيق طريقة

الأرجحية العظمى للتقدير، وذلك عندما يُجيب المفحوص عن جميع فقرات الاختبار إجابة صحيحة أو إجابة خطأ، وقد سُميّت هذه الطريقة بهذا الاسم لأنها تستخدم نظرية بييز في حساب الاحتمال المشروط والذي يتمثل بالمعادلة الآتية:

$$p(B/A) = \frac{P(A/B)P(B)}{P(A)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots[14]$$

(Kery,2010, p16)

إذ تُشير $P(A/B)$ إلى احتمالية حدوث الظاهرة (A) مع العلم أن الظاهرة (B) قد حدثت، وP(B) تُشير لاحتمالية حدوث الظاهرة (B)، وP(A) تُشير إلى احتمالية حدوث الظاهرة (A).

كما ويُمكن التعبير رياضيًا عن تقدير القدرة بطريقة بييز بواسطة المعادلة الآتية:

$$P(\theta/u) = \frac{P(u/\theta)P(\theta)}{P(u)}$$

(Fox,2010, p30)

حيث إن:

P (θ | u): التوزيع البعدي لـ (θ).

P(θ) : التوزيع القبلي لـ (θ).

u: تشير إلى نمط الإجابة .

θ: معلمة القدرة.

P(u | θ) : اقتران الأرجحية.

ويعُبر كل من هامبلتون وسوامنيثانز وروجرز(Hambleton,Swaminathan and Rogers,1991;p.38) عن معادلة التقدير البيزية بالصيغة الآتية:

$$f(\theta|u) \propto L(u|\theta)\, f(\theta)$$

حيث إن:

$f(\theta|u)$: هي التوزيع البعدي

L (u | θ) : اقتران الأرجحية ، f(θ) : التوزيع القبلي للقدرة.

واستنادًا للطريقة البيزية في تقدير القدرة جرى تطوير أكثر من أسلوب لتقدير قدرة المفحوص ومعالم الفقرات والتي تتمثل بالطرق الآتية:

أ) **أسلوب التقدير البعدي المتوقع (EAP)** Expected A Posterior:

في هذا التقدير لا يتم اســـتخدام التقدير الدوراني الذي يعتمد طريقة تقديرية تكرارية، وإنما يســتخدم طريقة مباشــرة؛ حيث يتم اســتخدام التوزيع الطبيعي المعياري وذلك بتقســيم قيم الســمة الكامنة التي تمثل هذا التوزيع والتي تقع بين(- ٣) و(٣) إلى (٦١) تجزيئًا بفترات طولها (صـفر وواحد) يُطلق على كل منها ربيعًا (Quadrant) ويُرمز له بالرمز (Qr)، كما يُعطى لكل من هذه الربيعات وزنٌ (Weight) يُرمز له بالرمز Wr(Qr) (Keller,2000).

يُستخدم الرمز L(Qr) للدلالة على الأس في اقتران لوغريتم الأرجحية عند كل من الربيعات (التي عددها في العادة ٦١ ربيعاً) الذي يُمكن اســـتخراجه من خلال اللوغريتم المقابل للوغريتم الأرجحية عند كل من الربيعات الذي يُمثل بالتالي المتغير العشوائي المرتبط بالوزن Wr (θ).

وتقدر قيمة (θ) في أسلوب توقع التوزيع البعدي من خلال العلاقة الآتية:

$$\hat{\theta} = \frac{\sum_{r=1}^{61}[Q_r \times L(Q_r) \times W(Q_r)]}{\sum_{r=1}^{61}[L(Q_r) \times W(Q_r)]} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots[15]$$

(Baker,2004; p87)

حيث إن:

Qr: قيم (θ) للعقد الربيعية.

W (Qr): الأوزان عند كل مربع.

L(Qr): لوغريتم الأرجحية العظمى للعقدة (log-likelihood).

ويمتاز هذا الأســلوب بأنه غير متكرر (Non-iterative) وســريع الحســاب، ويعطي تقديرًا للقدرة لكل أنماط الإجابات، أما جوانب الضـعف في هذا الأسـلوب فيتمثل في أنه عندما يكون عدد الفقرات قليلاً يصبح التقدير متحيزًا نحو وسط التوزيع، لأنه يعتمد على معلومات أولية عن وسط المجتمع وانحرافه المعياري(Hambleton , Swaminathan, and Rogers, 1991) .

ب) **أسلوب التقدير البعدي الأعظم** (MAP) Maximum A Posterior:

يقوم هذا الأسلوب على استخدام معلومات سابقة عن توزيعًات معلمة القدرة ومعالم الفقرات، الذي يتم افتراضــه من قبل الباحث بناءً على معلومات سـابقة لديه، ولعل التوزيع الأكثر اسـتخدامًا في هذا الســياق هو التوزيع الطبيعي الجرســي، أي أن ســمات المفحوصــين لها توزيع متوســطه (صفر) وانحرافه المعياري (واحد)، ويتم تقدير معلمة القدرة في هذا الأسلوب باستخدام خوارزمية

التوقع (Expect) والتعظيم (Maximum)، ففي مرحلة التوقع يتم حســـاب القيم المتوقعة للتكرارات على نقاط التربيع (Quadrature Points)، وبعد ذلك يتم تقديم القيم المتوقعة إلى معادلات التقدير لتقديرات (MAP) في مرحلة التعظيم، ويتم تكرار مرحلتي التوقع (E) والتعظيم (M) حتى يتقارب التقدير.

ويمتاز هذا الأســـلوب بالدقة في تقدير معلمة القدرة ومعالم الفقرات، كما ويمكن تقديرها (معلمة القدرة) لكل المفحوصـين، إلا أن تقديراته تتصـف بالتحيّز خاصـة عندما يكون عدد فقرات الاختبار أقل من (٢٠) فقرة (fox, 2010).

العوامل التي تؤثر في دقة تقدير قدرات الأفراد ومعالم الفقرات:

يرى هامبلتون وجونز (Hambleton and Jones, 1994) أن وجود أخطاء في تقدير معالم الفقرات وقدرات المفحوصـين ينعكس بشـكل سـلبي على دقة تقدير تلك المعالم كما يؤثرعلى دالة معلومات الاختبار، وعموماً تعتمد عملية تقدير معالم الفقرات وتقديرات معلمة القدرة للمفحوصـين على عوامل عدة، أهمها:

أ- اختيار الطريقة المناسـبة للتقدير، فقد أشـار بعض الباحثين إلى أن طريقة بييز تعمل بشـكل أفضــل من طريقة الأرجحية العظمى مع الاختبارات القصــيرة وبحجم عينة أقل (Gruijter and Kamp, 2002)

ب- تزداد الدقة في تقدير القدرة كلما قلّ الخطأ المعياري للتقدير (SEE) Standard Error of Estimate وفي هذا الســياق أشـــار كل من هامبلتون وســـوامينثـان وروجورز (Hambleton,Swaminathan and Rogers, 1991) إلى أن الخطأ المعياري للتقدير يعتمد على عوامل هي :

١- عدد فقرات الاختبار: فبزيادة عدد فقرات الاختبار تقل قيمة الخطأ المعياري في التقدير.

٢- القدرة التمييزية لفقرات الاختبار: فكلما زادت معلمة التمييز للفقرة قل الخطأ المعياري للتقدير؛ لأن إجابة الفقرة من خلال التخمين العشوائي يُصبح صعبًا.

٣- درجة المواءمة بين صعوبة الاختبار وقدرة المفحوصين: فكلما اقتربت صعوبة الفقرة من قدرة المفحوص قلّ الخطأ المعياري للتقدير ومن هنا فإن الاختبار السـهل جدًا أو الصـعب جدًا يرتبط بخطأ معياري كبير.

كما أشـــارت ســـتوكنج (Stocking,1990) إلى أن نجاح تطبيقات نظرية الاســـتجابة للفقرة كالقياس التكيفي، وبنوك الأسـئلة وغيرها من التطبيقات يعتمد بشـكل كبير على دقة التقديرات لكل من معالم الفقرات ومعلمة القدرة للمفحوصـــين، وفي هذا الصـــدد بينت ســـتوكنغ أنه من الممكن الحصـــول على تقديرات لمعالم الفقرات بشـــكل أدق، إذا كان التخطيط لاختيار عينة المعايرة يأخذ بعين الاعتبار دقة تقديرات معالم الفقرة، وقد قامت ستوكنج بإيضـاح العلاقة التي تربط بين معلمة قدرة المفحوص (θ) ودقـة تقدير معـالم للفقرات، من خلال اســـتخدام توقع المعلومـات، حيث استخدمت إجراءات الاشتقاق البسيط والاشتقاق الجزئي لدالة معلومات الفقرة بالنسبة لكل معلم من معالم الفقرة بهدف تحديد مســـتويات القدرة التي تقابل أقصـــى معلومات، عند تقدير معالم الفقرات للحصـول على دقة أكبر في تقدير تلك المعالم، وذلك من خلال استخدام المعادلة الآتية والتي قدمها لورد (Lord,1980)، حيث تبين أنه لأي معلم (x) تكون دالة المعلومات بالنســـبة لذلك المعلم على الصورة الآتية:

$$i_{xx} = \frac{1}{PQ}\left(\frac{\partial P}{\partial \chi}\right)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots[16]$$

(Stocking, 1990, P.463)

حيث إن:

i_{xx} : دالة معلومات المعلمة (x).

P: احتمال الإجابة الصحيحة.

Q :احتمال الإجابة الخطأ.

$\frac{\partial P}{\partial \chi}$: مشتقة الإجابة الصحيحة بالنسبة للمعلمة x.

وبالنســـبة لمعلمة التخمين فالمعادلة الآتية تُبين كمية المعلومات التي يُســـاهم بها المفحوص لتقدير معلمة التخمين (c).

$$i_{cc} = \frac{1}{(1-c)^2}\frac{Q}{P}\,. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots[17]$$

(Stocking, 1990, P.463)

حيث إن:

i_{cc}: تُمثل دالة معلومات الفقرة التي تبُين مدى مســـاهمة المفحوصـــين في تقديم معلومات لتقدير معلمة التخمين(c).

c: معلمة التخمين.

Q: احتمال الاجابة الخطأ.

P: احتمال الإجابة الصحيحة.

ولدى اشتقاق هذه المعادلة بالنسبة لمعلمة القدرة (θ) تنتج المعادلة الآتية:

$$\frac{\partial i_{cc}}{\partial \theta} = \frac{-Da}{(1-c)^3} \frac{1}{P^2} Q(P-c) = 0. \quad ....[18]$$

(Stocking,1990, P.462)

حيث إن:

$\frac{\partial i_{cc}}{\partial \theta}$: مشتقة دالة المعلومات بالنسبة لمعلمة القدرة (θ)

c: معلمة التخمين.

Q: احتمال الإجابة الخطأ.

P: احتمال الإجابة الصحيحة.

D : ثابت قيمته تساوي (١,٧٧).

وبمساواة المعادلة أعلاه بالصفر يُلاحظ أن قيمة المشتقة تساوي صفرًا عندما تكون قيمة (Q=0)؛ أي إن احتمال إجابة المفحوص عن الفقرة إجابة خطأ يساوي صفرًا؛ وهذا يحدث عندما تكون معلمة القدرة(θ=∞) وبالتالي تكون قيمة دالة المعلومات التي تُبنى من مساهمة المفحوص في تقدير معلمة التخمين مساوية للصفر (i_{cc} تساوي صفرًا).

كما تكون أيضاً قيمة (i_{cc}) مساوية للصفر عندما تكون احتمالية الإجابة الصحيحة تساوي معلمة التخمين (P=c)؛ أي ان احتمال الإجابة بدون تخمين يساوي صفرًا، وهذا يحصل عندما تكون معلمة القدرة للمفحوص تؤول الى سالب ما لا نهاية (θ=-∞) ، إذ تصبح قيمة دالة المعلومات التي

يُساهم بها هؤلاء الأفراد$i_{cc}=1/[(1-c)c]$ ، وبالتالي فإن المفحوصين ذوي القدرة المتدنية هم من يساهمون في إعطاء معلومات حول معلمة التخمين (c) وعندما تكون قيمة معلمة التخمين مساوية لـ (c=0.5)؛ فإن المفحوصين ذوي القدرة المتدنية هم من يساهمون بأقل مقدار من المعلومات في تقدير معلمة التخمين (c).

وبالتالي لا بد من المواءمة بين قدرة المفحوصين ودرجة صعوبة الاختبار المقدم لهم من أجل الحصول على تقديرات دقيقة لكل من معالم الفقرات ومعلمة القدرة للمفحوصين، وفي هذا السياق فقد أشار كل من هامبلتون وسوامينثان (Hambleton and Swaminthan, 1985) إلى أن الاختبار السهل من المتوقع أن يقدّم تقديرات أكثر دقة عند مستويات القدرة المتدنية ، فهو أكثر فائدة للمفحوصين ذوي القدرات المتدنية، أما الاختبار الصعب فمن المتوقع أن يُقدّم معلومات أكثر دقة عند مستويات القدرة العليا، وبالتالي فهو أكثر فائدة للمفحوصين ذوي القدرات العليا، ومن المتوقع أن يقدم الاختبار السهل معلومات أكثر فائدة وفعالية للمفحوصين ذوي القدرات المتدنية مقارنة مع الاختبار ذي المدى الواسع، كما أن الاختبار الصعب من المتوقع أن يقدّم معلومات أكثر فائدة وفعالية للمفحوصين ذوي القدرات العليا مقارنة مع الاختبار ذي المدى الواسع، وعند مستويات القدرة المتدنية فإن الاختبار ذا المدى الواسع، يحتاج إلى زيادة عدد فقراته بنسبة كبيرة ليصبح مساويًا للاختبار السهل في قياس القدرة، وكذلك عند مستويات القدرة العليا فإن الاختبار ذا المدى الواسع يحتاج إلى زيادة عدد فقراته بنسبة كبيرة ليصبح مساويًا للاختبار الصعب في قياس القدرة.

دالة معلومات الفقرة والاختبار.

تُعتبر دالة المعلومات من المفاهيم الأساسية في نظرية الاستجابة للفقرة سواء كانت الفقرة ثنائية التدريج أو متعددة التدريج؛ حيث ينظر لدالة معلومات الفقرة على إنها دالة للعلاقة بين قدرة المفحوص والمعلومات التي تقدمّها الفقرة، إذ تمثل هذه الدالة كمية المعلومات التي يُمكن أن توفرها فقرات الاختبار عند مستوى قدرة محدد، بالإضافة لإمكانية تحديد النقطة على متصل القدرة التي يمكن من خلالها الحصول على أقصى معلومات عند مستوى قدرة (θ) معين، ويُشير كل من هامبلتون وسوامنيثان وروجرز (Hambleton,Swaminathan and Rogers,1991) إلى أن دالة المعلومات تُعد وسيلة لوصف كل من الفقرات والاختبار، ولعل أهم التطبيقات التي تساهم بها دالة المعلومات هي:

أ- بناء الاختبارات المختلفة الاهداف والاغراض الخاصة بها واختيار فقرات الاختبار بناءً على كمية المعلومات التي تقدمها كل فقرة وبما يتلاءم مع طبيعة وهدف الاختبار.

ب- تقييم مدى دقة أدوات القياس، مما يتيح فرصة المقارنة بينها، والمقارنة بين طرق التصحيح المستخدمة فيها. إن ما يميز دالة المعلومات أنها يمكن أن تُعّرف عند كل مستوى من مستويات معلمة القدرة (θ) على متصل السمة، وأن دالة المعلومات لأي فقرة لا تتأثر بما تقدمه الفقرات الأخرى المُكونة للاختبار من معلومات (Lord,1980; De Ayala, 2009).

والمعادلة الآتية تبين طريقة حساب دالة معلومات الفقرة:

$$I_i(\theta) = \frac{P'(\theta)^2}{Pi(\theta)Qi(\theta)} \quad ....[19]$$

حيث إن:

$I_i(\theta)$: دالة المعلومات للفقرة (i).

$P'(\theta)$: المشتقة الأولى لاحتمال الإجابة الصحيحة على الفقرة (i).

$Pi(\theta)$: احتمال الإجابة الصحيحة على الفقرة (i).

$Qi(\theta)$: احتمال الإجابة الخطأ على الفقرة (i).

(Yen, and Fitzpatrick, 2006 ,P 126)

أثر معالم الفقرة على دالة المعلومات:

يمكن تحديد الأثر الذي تتركه معالم الفقرات في إطار النموذج اللوجستي ثلاثي المعلمة على دالة المعلومات من خلال المعادلة الآتية:

$$I(\theta,Ui)max = \frac{D^2 ai^2}{8(1-ci^2)}[1 - 20ci - 8ci^2 + (1+8ci)^{3/2}] \quad ....[20]$$

(Hambleton,Swaminathan, 1985, p105)

حيث إن:

$I(\theta, Ui)_{max}$: القيمة القصوى لدالة المعلومات، c_i : تشير لمعلمة التخمين.

a_i : معلمة التمييز.

D: ثابت ويساوي (١,٧٧).

إذ يُلاحظ من المعادلة أعلاه أن هناك اثرًا لكل من معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين على دالة المعلومات؛ فالمعلومات تزداد كلما كانت قيمة معلمة الصعوبة (b) قريبة من مستوى القدرة (θ) للمفحوص، كما تزداد قيمة دالة المعلومات باقتراب معلمة التخمين (c) من الصفر، حيث يمكن الحصول على أعلى قيمة لدالة المعلومات حينما تصبح قيمة معلمة التخمين مساوية للصفر، وعندها فإن القيمة القصوى لدالة المعلومات (θ_{max}) تساوي معلمة الصعوبة (b)، وعندما تكون قيمة معلمة التخمين أكبر من صفر ($c > 0$) فإن قيمة (θ_{max}) تُصبح إلى يمين معلمة الصعوبة (b) وبازدياد قيمة معلمة التمييز (a) يزداد ميل منحنى الفقرة، ويقل التباين وترتفع قيمة المعلومات (Hambleton,Swaminathan and Rogers,1991).

دالة معلومات الاختبار:

أشار كل من هامبلتون وجونز (Hambleton and Jones,1994) إلى أن دالة المعلومات التي يمكن أن يوفرّها الاختبار عند مستوى قدرة (θ) ما هي إلا مجموع دوال المعلومات للفقرات المكونة لهذا الاختبار عند ذلك المستوى للقدرة (θ).

والمعادلة الآتية تُبين قيمة دالة المعلومات للاختبار:

$$I(\theta) = \sum_{i=1}^{n} I_i \ (\theta) \qquad ....[21]$$

(Hambleton,1989, P.164)

حيث إن:

$I_i(\theta)$: دالة معلومات الفقرة.

$I(\theta)$: دالة معلومات الاختبار .

تمتاز دالة معلومات الاختبار عموما باستقلاليتها عن عينة المفحوصين التي حسبت بناءً عليها، وبذلك تقدم النظرية الحديثة في القياس مميزات إضافية، فيما يتعلق بزيادة القدرة على تقدير أخطاء القياس. إن مقدار المعلومات التي يزوّد بها الاختبار عند مستوى قدرة (θ) يرتبط عكسيًا مع مستوى الدقة إذ توضح ذلك المعادلة الآتية:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \qquad ....[22]$$

(Hambleton,1989, P.164)

حيث إن:

SE(θ) : هي الخطأ المعياري للتقدير Standard Error of Estimate

I(θ) : دالة المعلومات.

وبالتالي فإن معدل الخطأ المعياري للتقدير Standard Error of Estimate (SEE) يرتبط مع الثبات المعروف وفق النظرية الكلاسيكية بالعلاقة $r_{xx}=1-(SEE)^2$ ، إذ تُبيّن هذه العلاقة أن الاختبار ذا الثبات المرتفع تكون قيمة الخطأ المعياري للتقدير له متدنية وبالتالي فإن كمية المعلومات التي يقدمها تكون كبيرة وذلك لأن الخطأ المعياري للتقدير يرتبط عكسيًا مع الجذر التربيعي لدالة معلومات الاختبار(Linden and Hambleton,1997) .

المحكات المستخدمة للحكم على دقة تقدير معالم الفقرة ومعلمة القدرة:

تتطلب عملية تقدير معالم الفقرات توخي الدقة بحيث يكون خطأ التقدير لها أقل ما يمكن، إذ أشـــار لورد(Lord,1980) في هذا الســياق إلى مجموعة من المحكات التي ينبغي الاســتناد إليها في الحكم على دقة التقدير وتقييم مدى جودة الأسلوب الإحصائي المُستخدم للوصول إلى قيم تقديرية للمعلم وهي:

أ- **التحيّز** (Bias)

يُعبر عن عدم التحيز بالدرجة التي تكون فيها القيم المقدرة لمعالم الفقرة وفق الطريقة المســتخدمة قريبة من القيم الحقيقية لتلك المعالم، وكلما اقتربت قيمة هذا المؤشر من الصفر كلما دلّ ذلك على دقة أعلى، إذ يتم إيجاد مؤشـــر عدم التحيز عن طريق حســـاب الفرق بين القيم المقدرة والقيم الحقيقية وفق المعادلة الآتية:

$$\text{bias} = \frac{\sum_{j=1}^{N}(\hat{\theta}_j - \theta_j)}{N} \qquad \text{.......................[23]}$$

(Weiss a

حيث إن:

$\hat{\theta}_j$: تمثل المعلم المقدر.

θj:القيمة الحقيقية للمعلم.

J : العدد الكلي للفقرات.

وكلما اقتربت قيمة مؤشر التحيز من الصفر كان ذلك مؤشرًا على الدقة في تقدير معالم الفقرات ومعلمة القدرة للمفحوصين.

ب- **الجذر التربيعي لمتوسط مربعات الخطأ** Root Mean Squared Error (RMSE)

يُقاس هذا المؤشـــر من خلال إيجاد الجذر التربيعي لمربع الفرق بين القيمة المقدرة والقيمة الحقيقية، لكل من معالم الفقرات أو معلمة القدرة للمفحوصين، وكلما اقتربت قيمة هذا المؤشر من الصفر دلّ ذلك على دقة التقدير، ويتم إيجاده من خلال الصيغة الرياضية الآتية:

$$RMSE_x = \sqrt{\frac{\sum_{i=1}^{n} (\hat{X}_{estmated} - X_{true})^2}{n}} \quad ....[24]$$

(Natesan and Kieftenbeld, 2012; p405)

حيث إن:

n : حجم العينة

$_i$: القيمة الأولى للمعلم.

$X_{estimated}$: قيمة المعلم المقدرة.

X_{true} : القيم الحقيقية للمعلم.

الكفاءة النسبية (Relative Efficiency)

أشـــــار كل من ين و فتزباترك (Yen and Fitzpatrick, 2006) ان لدالة المعلومات أمكانية المقارنة بين الاختبارات المختلفة التي تقيس القدرة نفســـها، أو تحري فعالية نماذج مختلفة اســـتخدمت لتقدير معالم الفقرات أو معلمة القدرة للمفحوصـــين، أو طرق مختلفة لتصـــحيح الاختبار ذاته، أو مقارنة اســـتراتيجيات مختلفة اســـتخدمت لبناء الاختبارات مثل المقارنة بين الاختبارات التقليدية والاختبارات التكيفية، ويُعّبر عن مؤشر الكفاءة النسبية من خلال الصـيغة الرياضية الآتية:

$$RE(\theta) = \frac{I_A(\theta)}{I_B(\theta)} \qquad ....[25]$$

(Yen and Fitzpatrick, 2006 ;P.129)

حيث إن:

$RE(\theta)$: الكفاءة النسبية

$I_A(\theta)$: دالة معلومات الاختبار (A)

$I_B(\theta)$: دالة المعلومات للاختبار (B).

الدراسات السابقة

حظيت قضية معالجة التخمين في الأسئلة الموضوعية/الاختيار من متعدد بقدرٍ لا بأس به من الاهتمام من الباحثين في إطار كلٍ من النظرية الكلاسيكية في القياس ونظرية الاستجابة للفقرة، ويتناول هذا الفصل مجموعة من الدراسات والأبحاث التي تناولت إجراءات ونماذج معالجة التخمين في الأسئلة الموضوعية/الاختيار من متعدد وأثرها على دقة تقدير معالم الفقرات ومعلمة القدرة.

لعلّ المُراجع للأدب التجريبي الذي تناول موضوع التخمين في إطار الأسئلة الموضوعية الاختيار من متعدد يُلاحظ التنوع في الدراسات التي أجريت في إطار نظرية الاستجابة للفقرة والتي تناولت معالجة أثر التخمين في الاختبارات الموضوعية/الاختيار من متعدد من حيث طبيعة البيانات التي تم إجراء المقارنة بناءً عليها، حيث تناولت بعض هذه الدراسات بيانات افتراضية قائمة على المحاكاة، تم توليدها باستخدام برمجيات حاسوبية مخصصة لذلك، في حين استخدمت دراسات أخرى بيانات حقيقية مستمدة من تطبيق مجموعات مختلفة من الفقرات على عينات محددة من المفحوصين، مثل الاختبارات الدولية في الرياضيات والعلوم، أو اختبارات من إعداد الباحث، وهناك بعض الدراسات استخدمت بيانات افتراضية(بيانات مولّدة) وبيانات حقيقية في نفس الوقت، كما اختلفت النماذج اللوجستية التي تعاملت معها الدراسات السابقة حيث استخدمت بعضها نماذج لوجستية مختلطة ونماذج معدلة إضافة إلى نماذج نظرية الاستجابة للفقرة التقليدية.

كما ويوجد أيضًا بطبيعة الحال تنوعٌ بالدراسات التي عالجت التخمين في إطار النظرية الكلاسيكية في القياس من ناحية الطرق والإجراءات التي استخدمتها وقارنت بينها، فالأدب التجريبي في إطار النظرية الكلاسيكية في القياس والمتعلق بمعالجة التخمين يزخر بكثيرٍ من هذه الدراسات، ويُلاحظ أن جميع الدراسات التي عالجت التخمين في إطار النظرية الكلاسيكية في القياس استخدمت بيانات حقيقية مستمدة من تطبيق اختبارات موضوعية/ اختيار من متعدد على عينات محددة من المفحوصين.

ولتسهيل استعراض الدراسات السابقة، بشكل يخدم أغراض هذه الدراسة، تم تقسيمها إلى قسمين على النحو الآتي:

اولاً: الدراسات التي سعت لمعالجة التخمين في إطار النظرية الكلاسيكية في القياس:

هنالك العديد من الدراسات التي هدفت إلى فحص فاعلية عدد من الطرق لمعالجة التخمين في إطار النظرية الكلاسيكية في القياس، ففي دراسة أجراها النبهان (Alnabhan,2002) دراسة هدفت إلى تقصّي أثر ثلاث طرق للتصحيح ضد أثر التخمين وهي: (عدد الإجابات الصحيحة، ومعاقبة التخمين، ومكافأة التخمين الجزئي) على الخصائص السيكومترية (الثبات و الصدق) للاختبار، حيث تكّونت عينة الدراسة من (١٢٠) طالباً وطالبة في مرحلة ما قبل التخرج في جامعة أبوظبي في دولة الإمارات العربية المتحدة، وتألّف اختبار الدراسة من (٤٠) فقرة اختيار من متعدد بواقع أربعة بدائل لكل فقرة في مساق منهجية البحث النفسي، كما استخدم الباحث اختباراً قصيرًا مكونًا من (١٠) فقرات للمحتوى التعليمي نفسه الذي يغطيه الاختبار المكون من (٤٠) فقرة وببدائل جميعها خاطئة لكل فقرة؛ من أجل تصنيف المفحوصين إلى: ذوي مخاطرة مرتفعة وذوي مخاطرة متدنية، فالطالب الذي أجاب عن عدد اكبر من الفقرات اعُتبر ذا مخاطرة مرتفعة والذي أجاب عن عدد أقل من الفقرات أعتبر ذا مخاطرة منخفضة، وقد أظهرت نتائج الدراسة أن الطرق الثلاث المختلفة أعطت معاملات صدق وثبات مختلفة عن طريقة المعرفة الجزئية -التي تقوم على افتراض أن المفحوص يمتلك معرفة تمكنه من حذف بعض البدائل عند اجابته عن الفقرة-، حيث كان معامل الثبات للاختبار المصحح ضد أثر التخمين باستخدام طريقة المعرفة الجزئية (٠,٨١) بينما كان (٠,٦٥) عند استخدام طريقة التصحيح باستخدام عدد الإجابات الصحيحة، كما بلغ (٠,٧٣) عند استخدام طريقة معاقبة التخمين، وأكدت الدراسة أن التخمين يُقلل من صدق وثبات الاختبار من خلال إعطاء تباين غير حقيقي في نتائج الاختبار.

كما أجرى الخرشة(٢٠٠٤) هدفت إلى المقارنة بين ثلاث طرق(المكأفاة، والعقاب و المكافأة والعقاب معاً) للتصحيح ضد أثر التخمين في اختبارات الاختيار من متعدد، حيث قام الباحث بإعداد اختبار تحصيلي مكون من (٦٠) فقرة من نوع الاختيار من متعدد، بواقع أربعة بدائل للفقرة الواحدة بهدف قياس المصطلحات الأساسية في مقرر مبادئ القياس والتقويم، وتكونت عينة الدراسة من (١٩٧) طالبًا وطالبة في كلية التربية في جامعة مؤته: في العام الدراسي (٢٠٠٣/٢٠٠٤م) وللحكم على فاعلية الطرق المُستخدمة قام الباحث باستخدام المؤشرات الإحصائية التالية: اختبار (ت) للعينات المستقلة، ومعاملات الثبات، ومقاييس النزعة المركزية، ومقاييس التشتت، والخطأ المعياري للقياس لعلامات المفحوصين وفق كل طريقة من طرق التصحيح لأثر التخمين، وفي ضوء نتائج الإحصاءات المستخرجة أوصت الدراسة باستخدام معادلة التصحيح من أثر التخمين (العقاب والمكافأة) في تصحيح اختبارات الاختيار من متعدد.

وفي دراسة أجراها الزبون (2004) هدفت إلى المقارنة بين أثر ثلاث طرق لتصحيح الإجابات على الاختبارات الموضوعية/ الاختيار من متعدد لضبط التخمين على الخصائص السيكومترية للاختبار والفقرة في ضوء نظرية الاستجابة للفقرة حيث جرى استخدام الطريقة التقليدية ومعادلة التصحيح لأثر التخمين وطريقة الاختيار الجزئي، حيث تمّ إعداد ثلاث صور متكافئة لاختبار تحصيلي لطلبة مساق (الإحصاء التربوي) في جامعة مؤتة، وجرى تطبيقها على عينة مكونة من (٤٠٦) طلاب في كلية التربية في جامعة مؤتة، جرى تصحيح الإجابات على كل صورة من صور الاختبار بإحدى طرق التصحيح الثلاث، ثم حُللّت البيانات باستخدام برمجية ونستب (Winstep) وجرى تقدير مَعلمة القدرة (θ) والخطأ المعياري في القياس لقدرة كل فرد، كما تم استخراج مؤشرات على معاملات الثبات لتباعد الأفراد، ومعاملات صدق المحك التلازمي، ومعاملات صعوبة الفقرة، ومعاملات تمييز الفقرة.

وأظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين معاملات الثبات لصالح استخدام كل من طريقة (التصحيح لأثر التخمين) وطريقة (الاختيار الجزئي)، كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية بين معاملات صدق الاختبار لصالح الطريقة التقليدية وطريقة الاختيار الجزئي، إضافة إلى ذلك أظهرت الدراسة أن هناك فروقًا ذات دلالة إحصائية بين معاملات صعوبة الفقرة، لصالح معادلة التصحيح لأثر التخمين، وأن هناك فروقًا ذات دلالة إحصائية بين معاملات تمييز الفقرة عند استخدام الطريقة التقليدية وطريقة (الاختيار الجزئي).

أما راشد(٢٠١١) فقد أجرى دراسة هدفت إلى معرفة أثر التعليمات الخاصة بالتخمين في استجابة الطلبة لفقرات اختبار موضوعي/ اختيار من متعدد على درجاتهم وعلى الخصائص السيكومترية للاختبار(الصدق والثبات) وعلى خصائص الفقرات (الصعوبة والتمييز) ولتحقيق أهداف دراسته قام ببناء اختبار تحصيلي في وحدتين دراسيتين من مادة الحضارة العربية الإسلامية، حيث تكّون الاختبار من (٤٠) فقرة من نوع الاختيار من متعدد، بأربعة بدائل لكل فقرة، طُبّق الاختبار على عينة مكونة من (٦٦٦) طالبا وطالبة من الملتزمين وغير الملتزمين بتعليمات التخمين من كلية التربية الأساسية في جامعة الكويت، وأظهرت نتائج الدراسة وجود فروق دالة إحصائيًا بين درجات التحصيل للمجموعتين وذلك لصالح مجموعة الملتزمين بتعليمات التخمين حيث بلغ المتوسط الحسابي لدرجاتهم(٠,٦٦) في حين كان المتوسط الحسابي لدرجات الطلبة غير الملتزمين بتعليمات التخمين (٠,٤٠).

ثانيا: الدراسات التي سعت لمعالجة التخمين في إطار نظرية الاستجابة للفقرة:

أجرى والر (Waller, 1989) دارسة هدفت إلى المقارنة بين درجة مطابقة البيانات للنماذج اللوجستية الآتية: النموذج اللوجستي الثلاثي المعّلمة ونموذج إزالة التخمين العشوائي المستند إلى القدرة (الذي يفترض أن المفحوص يخمن فقط على الفقرات المرتفعة الصعوبة بالنسبة له، حيث يقوم هذا النموذج بحذف اجابات المفحوصين على تلك الفقرات)، والنموذج اللوجستي الثلاثي المعلمة مع تثبيت معلمة التخمين. تكونت عينة الدراسة من ثلاث مجموعات من الطلبة بما يقارب (٣٠٠٠) مفحوص من الصفوف الأساسية والصفوف المتوسطة والثانوية في الولايات المتحدة الامريكية، تقدموا جميعا لبطارية اختبارات تكونت من(٣٣)اختبارًا، تقيس المهارات الأساسية الشاملة (CTBS) Comprehensive Tests of Basic Skills من المرحلة الأساسية إلى المرحلة الثانوية في مجالات: المفردات، والاستيعاب، والتعبير الكتابي، والمهارات الرياضية، واستخدم الباحث برمجية (Logog) لإيجاد تقديرات معالم الفقرات، وقدرات المفحوصين باستخدام شبه طريقة الأرجحية الهامشية، وهو إجراء لا يتضمن لوغريتم التوقع والتعظيم (EM) Expect and Maximize؛ بل يفترض هذا الاجراء توزيعًا محدداً لمعلمة القدرة وغالبا ما يكون التوزيع الطبيعي.

كما واستخدم الباحث الإحصائي كاي سكوير (χ^2) للتحقق من حسن المطابقة في النماذج المستخدمة في الدراسة وهي: النموذج اللوجستي الثلاثي المعلمة، والنموذج اللوجستي الثلاثي المعلمة مع وجود معلمة تخمين ثابتة قيمتها تساوي (1/K) حيث(K) هي عدد البدائل، ونموذج Ability Removing Random Guessing (ARRG) وبالرغم من أن كلاً من النماذج السابقة تهدف إلى إزالة أثر التخمين العشوائي، إلا أن نتائج الدراسة أظهرت أن كلاً من النموذج اللوجستي الثلاثي المعلمة ونموذج (ARRG) كانت مطابقتهما للبيانات أفضل من النموذج اللوجستي الثلاثي المعلمة ذي التخمين الثابت، كما أشارت نتائج الدراسة إلى أن النموذج اللوجستي الثلاثي المعلّمة أفضل من نموذج (ARRG) في دقة تقديرات القدرة ومعالم الفقرات، وعُزي ذلك إلى أن النموذج الثلاثي المعلمة بإمكانه معالجة المعرفة الجزئية، والتخمين العشوائي، واستنتج أن معلمة التخمين يجب أن تُقدّر بشكل منفرد وليس بشكل ثابت لجميع فقرات الاختبار.

وفي دراسة أجراها ستيدي (Setiadi, 1997) قارن خلالها بين النموذج اللوجستي الأحادي المعلمة المعدّل عند تثبيت معلمة التخمين (c= 0.20) مع كل من النموذج اللوجستي الأحادي المعلمة والنموذج اللوجستي الثلاثي المعلمة، واستخدم الباحث برمجية تستجراف (Test Graf) لتقدير معلمة القدرة للمفحوص ومعالم الفقرات(معلمة التخمين ومعلمة التمييز ومعلمة الصعوبة) للنموذج اللوجستي

الثلاثي المعلمة، كما استخدم برمجية (Bilog) لاستخراج تقديرات معالم الفقرات ومعلمة القدرة للمفحوص للنموذج اللوجستي الأحادي المعلمة، وتمّ الاعتماد على منهج توليد البيانات بناءً على النموذج اللوجستي الثلاثي المعلمة، إذ وُلدت البيانات لاختبارين من نوع الاختيار من متعدد، الأول مكون من (٣٠) فقرة، والثاني مكون من (٦٠) فقرة وبلغ حجم العينات كما يأتي:

(١٠٠ مفحوص و ٢٠٠ مفحوص و ٥٠٠ مفحوص) على التوالي، وتم تقييم النتائج المستخرجة في ضوء المحكات الآتية: قيم معاملات الارتباط بين القيم المُقدرة وتلك المولدة، والتحيز، والجذر التربيعي لمتوسط مربعات الأخطاء لمعالم الفقرات ولمعلمة القدرة للمفحوص، وأشارت نتائج الدراسة إلى أن النموذج اللوجستي الأحادي المعلمة أظهر تقديرات لمعلمة القدرة أدق من النماذج الأخرى المشمولة في الدراسة في حالة توزيع معلمة القدرة توزيعًا طبيعياً، وحينما كان توزيع البيانات توزيعًا منتظماً (Uniform)، بينما أظهر النموذج اللوجستي الثلاثي المعلمة المُعدّل دقة أكبر في تقدير معالم الفقرات مقارنة بالنماذج الأخرى في الدراسة.

كما أجرى بيلتون (Pelton, 2002) دراسة هدفت إلى مقارنة الدقة والاستقرار في تقدير معلمة الصعوبة، ومعلمة القدرة باستخدام النظرية الكلاسيكية في القياس، والنماذج اللوجستية في نظرية الاستجابة للفقرة، باستخدام بيانات مولّدة (استخدام أسلوب المحاكاة)، واستخدم الباحث عدة برمجيات للإجابة عن أسئلة دراسته: برمجية الإكسل للإجابة عن الأسئلة المتعلقة بالنظرية الكلاسيكية، وكل من برمجية (WINSTEP) و (BIGSTEPS) لنموذج راش، وبرمجية (BILOG) للنموذجين اللوجستيين الثنائي المعلمة والثلاثي المعلمة، وتوصلت الدراسة إلى مجموعة من النتائج منها: أن تقديرات معلمة القدرة يمكن مقارنتها عبر النظرية الكلاسيكية ونظرية الاستجابة للفقرة، حيث تتباين تقديرات القدرة حسب كمية المعلومات المتوفرة في مجموعة البيانات المولدة، التي تتأثر بدورها بأحادية البعد، ودرجة التخمين ومدى التباين في صعوبة الفقرات مقارنة بقدرات الأفراد، كما توصلت الدراسة إلى أنه بوجود حجم عينة معتدل (٩٩٩) مفحوصًا وعدد فقرات ملائم (٣٣) فقرة، فإن النموذج اللوجستي الثنائي المعلمة يُقدم تقديرات أكثر دقة لمعلمة الصعوبة من نموذج راش والنموذج اللوجستي الثلاثي المعلمة والنظرية الكلاسيكية في القياس وذلك في ظل وجود تخمين قليل في الإجابة على فقرات الاختبار.

وأجرى بالدوين (Baldwin, 2006) دراسة هدفت إلى المقارنة بين كل من النموذج اللوجستي أحادي المعلمة، والنموذج اللوجستي ثنائي المعلمة من جهة والنموذج اللوجستي الثنائي المعلمة المعدّل من جهة أخرى، سعياً لمعرفة أثر النموذج الثنائي المعلمة المُعدل في تحسين تقديرات معالم الفقرات، استخدم الباحث نوعين من البيانات: مولّدة وحقيقية، وبأحجام مختلفة بلغ

مجموعها(١٣٥,٠٠٠) مفحوص في اختبارات تحصيلية في مبحث الرياضيات، وتألفت الاختبارات من (٢٥ فقرة و ٤٠ فقرة و ٥٠ فقرة) على التوالي، واستخدم الباحث برنامج (Multilog)لتقدير معالم الفقرات وقدرات المفحوصين، وأظهرت نتائج الدراسة أن النموذج الثنائي المعدل يُقللّ من أخطاء القياس مقارنة بالنموذج الأحادي المعلمة والنموذج الثنائي المعلمة، وبالنسبة لدقة تقديرات معالم الصعوبة فقد أظهرت النتائج أن النموذج الثنائي المعدل أدق في التقدير من النموذج الثنائي والأحادي في كل الحالات عدا حالة واحدة عندما حجم العينة (١٠٠) مفحوص، وطول الاختبار (٤٠) فقرة، كما بين بالدوين أن النتائج على البيانات الحقيقية والمولدة جاءت متماثلة.

كما أجرى كل من وايز وديمارز (Wise and DeMars, 2006) دراسة هدفت إلى مقارنة دقة تقديرات معالم الفقرات ودالة المعلومات وأثر تغير نسبة اللجوء إلى سلوك التخمين من قبل المفحوص على ذلك، حيث تم استخدام: النموذج المعدل المعتمد على الزمن والنموذج اللوجستي الثنائي المعلمة، وجرى استخدام بيانات مولّدة بنسب تخمين (١٠% و٣٠% و ٥٠%) واستخدمت ثلاثة أحجام عينات بواقع (٥٠٠ مفحوص) لكل عينة، وحُلّلت البيانات حسب النموذج اللوجستي الثلاثي المعلمة أولاً، وحسب النموذج المعدّل المعتمد على الزمن، أظهرت نتائج الدراسة أن التحّيز في تقدير معالم الفقرات ومعلمة القدرة أقل في حالة استخدام النموذج المعدّل المعتمد على الزمن، وكلما ارتفعت نسبة التخمين ارتفعت نسبة التحيّز في النموذج اللوجستي الثلاثي المعلمة، كما أشارت نتائج الدراسة إلى أن أخطاء التقدير أقل ودالة معلومات الاختبار أفضل في حالة النموذج المعدل المعتمد على الزمن.

كما أجرى عبابنة (٢٠٠٤) دراسة هدفت لتفحّص أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة ومعلمة القدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة، حيث تم إعداد اختبار قدرة عقلية مكون من (٧١) فقرة، وتكونت عينة الدراسة من (١٠٠٠) طالب وطالبة، واستخدم الباحث برمجية (Bilog 3.11) لتقدير قدرات المفحوصين ومعالم فقرات الاختبار، والأخطاء المعيارية في التقدير وإحصائيات مطابقة البيانات للنموذج اللوجستي الثلاثي المعلمة، وقد توصّلت الدراسة إلى أن الدقة في تقدير معالم الفقرة تزداد بزيادة حجم عينة المفحوصين وبزيادة تباين قدرة المفحوصين، كما تزداد الدقة في تقدير معلمة الصعوبة ومعلمة القدرة عندما يكون مدى القدرة للمفحوصين متوافقًا مع صعوبة الفقرات، وكذلك تزداد الدقة في تقدير معلمة التخمين عند استخدام عينة من ذوي القدرة المتدنية في معايرة الفقرات،

كما وتتسم تقديرات معالم الفقرة بعدم الاستقرار عند استخدام عينات متغايرة القدرة، وتكون تقديرات معلم القدرة أكثر دقة عند استخدام الطريقة المعيارية في اختيار الفقرات مقارنة بالطريقة العشوائية.

وفي دراسة أجراها كونج (Kong, 2007) لمعرفة أثر التخمين السريع على تقدير كل من معالم الفقرة ومعلمة القدرة للمفحوص وذلك من خلال قياس الزمن المستغرق في الإجابة ومقارنته مع الزمن الفعلي اللازم للإجابة عن الفقرة إجابة صحيحة، فإذا قلّ الزمن الذي يستغرقه المفحوص للإجابة عن تلك القيمة المحددة فإن المفحوص يكون قد لجأ إلى التخمين السريع، واستخدم الباحث النماذج التالية: نموذج توسط الجهد الثنائي المعلمة (Effort-Moderated Model) ونموذج توسط الجهد الثلاثي المعلمة، ونموذج توسط الجهد الأحادي المعلمة، ونماذج نظرية الاستجابة للفقرة الثنائية التدريج: النموذج اللوجستي الثلاثي المعلمة، والنموذج اللوجستي الثنائي المعلمة، والنموذج اللوجستي الأحادي المعلمة، وبلغت عينة الدراسة(٦٢٠) طالبا وطالبة من مستوى الدراسة الجامعية، حيث تقدموا لاختبار مكون من (٤٢) فقرة اختيار من متعدد، وجرى مقارنة نتائج هذه النماذج باستخدام المحكات الآتية: معاملات الارتباط، والتحيز، والجذر التربيعي لمتوسط مربع الخطأ.

أظهرت نتائج الدراسة أن مقدارًا بسيطًا من التخمين السريع يمكن أن يجعل تقديرات كل من معالم الفقرة ومعلمة القدرة للمفحوصين متحيزة إلى درجة كبيرة ويزداد مقدار هذا التحيز بزيادة نسبة التخمين السريع من قبل الطلبة، كما بينت نتائج هذه الدراسة أن نماذج توسط الجهد الثلاثة أعطت تقديرات أدق لمعالم الفقرات ومعلمة القدرة للمفحوصين ولدالة معلومات الاختبار مقارنة بنماذج نظرية الاستجابة للفقرة التقليدية(النماذج اللوجتسية :الأحادي المعلمة والثنائي المعلمة والثلاثي المعلمة).

وأجرى ستيزر (Stetzer ,2007) دراسة هدفت لتحرّي أثر التخمين السريع على تقديرات كل من معلمة القدرة للمفحوص ومعلمة الصعوبة للفقرة، مستخدماً نموذج توسط الجهد المعدّل (Effort-Moderated IRT Model) وقد استخدم الباحث عينة مكونة من (١٠,٤٠١) مفحوص من طلبة السنة الأولى في الجامعة، كما واستُخدم اختبار مُطّور من قبل مركز خدمات الاختبارات التربوية يقيس معرفة المفحوص بالنظريات والمفاهيم والحقائق في ميدان تخصصه، تكوّن الاختبار من (١٢٠) فقرة موضوعية من نوع اختيار من متعدد، تم تطبيقه على مرحلتين، تمثلت التعليمات المقدمة للمفحوصين بضرورة الإجابة عن جميع فقرات، ممّا ضمن عدم وجود بيانات مفقودة، وقد جرى تقديم الاختبار من خلال الحاسوب، حيث تم تسجيل زمن الاستجابة لكل مفحوص على كل فقرة بالثواني، وحُسبت عتبة الوصول للحل لكل فقرة والتي تُمثل الحد الفاصل بين سلوك الحل

وسلوك التخمين من قبل المفحوص تجاه الفقرة، حيث جرى بعدها تدريج أو معايرة البيانات باستخدام النموذج اللوجستي الثلاثي المعلمة، ونموذج توسط الجهد باستخدام برمجية (Bilog).

وأظهر نموذج توسط الجهد مطابقة للبيانات أفضل من النموذج اللوجستي الثلاثي المعلمة، أما بالنسبة لتقديرات المعالم: الصعوبة والتمييز والتخمين، فقد أظهرت نتائج الدراسة أن تقديرات معلمة الصعوبة ومعلمة التمييز كانت متشابهة بشكل كبير لكلا النموذجين، وكان معامل الارتباط بين هذه المعالم (٠,٩٩) أما بخصوص معامل الارتباط بين تقديرات معلمة التخمين في كلا النموذجين فقد بلغت (٠,٩٢).

أظهرت نتائج الدراسة وجود أربع فقرات كانت قيم معلمة التخمين لها مساوية للصفر عند تقديرها باستخدام نموذج توسط الجهد المعدّل، في حين كانت قيم معلمة التخمين لنفس الفقرات في النموذج اللوجستي الثلاثي المعلمة أعلى من القيمة (صفر)، كذلك أشارت النتائج إلى أن قيم دالة المعلومات كانت متطابقة عند جميع مستويات معلمة القدرة للمفحوص في كلا النموذجين، وكان الاختلاف بينهما عندما تكون قدرة المفحوص مساوية للصفر، وعند مستوى القدرة ٢ فما فوق.

وفي دراسة اجراها كل من بارشال وكروموري وشيزون ويي (Parshall, Kromery, Chason and Yi, 2007) حيث هدفت الدراسة إلى المقارنة بين النموذج اللوجستي الثلاثي المعلمة والنموذج اللوجستي الثنائي المعلمة والاحادي المعلمة وبين ثلاثة نماذج لوجستية معدلة من حيث دقة تقديرات معالم الفقرات ومعلمة القدرة واثر حجم عينة المفحوصين على تلك التقديرات.

وفي دراسة قام بها سوبادي (Subedi, 2009) بهدف تحري العلاقة بين التخمين الناجح والسمات الكامنة في نماذج استجابة الفقرة الثنائية، حيث استخدم الباحث في دراسته النماذج الآتية: النموذج اللوجستي ثنائي المعلمة، والنموذج اللوجستي الثلاثي المعلمة، ونموذج (MixIRT-R) (نموذج استجابة الفقرة المختلط للتخمين العشوائي) الذي يفترض أن سلوك التخمين الذي يقوم به الطلبة هو تخمين عشوائي، ونموذج (MixIRT-A) (نموذج استجابة الفقرة المختلط للتخمين المبني على القدرة) حيث يفترض هذا النموذج أن التخمين ليس دائما عشوائيًا وإنما هو مبني على قدرة المفحوص، واستخدم الباحث نظام التحليل الاحصائي(SAS) لتوليد البيانات، إذ تمّ توليد بيانات لاختبارين بعدد فقرات (٢٥ فقرة و ٥٠ فقرة) ، وأحجام عينات مختلفة حيث بلغت (٥٠٠ مفحوص و٢٠٠٠ مفحوص) على التوالي، واستخدم الباحث برنامج (Winbug) لتحليل البيانات، واستخدم الباحث المحكات الآتية: مربع مجموع الانحرافات للخطأ والتحيز ومعاملات الارتباط بين القيم المولّدة (الحقيقية) وتلك المقدرة من خلال برنامج (WinBug) للمفاضلة بين النماذج الأربعة التي

اشتملت عليها هذه الدراسة، خلصت هذه الدراسة إلى أن النماذج المختلطة المستخدمة أعطت تقديرات أدق لكل من معالم الفقرة ومعلمة القدرة للمفحوصين مقارنة مع كل من النموذج اللوجستي ثنائي المعلمة والنموذج اللوجستي ثلاثي المعلمة، كما وأوصت الدراسة باستخدام النماذج المختلطة التي تم اقتراحها في الاختبارات واسعة النطاق، وبنوك الأسئلة والاختبارات التكيفية.

وقام فو(Fu,2010) بدراسة سعت إلى مقارنة الدقة في تقدير معلمة قدرة المفحوصين، ومعلمة الصعوبة للفقرة باستخدام نماذج نظرية الاستجابة للفقرة، في ظل الظروف التالية: وجود التخمين على فقرات الاختبار الموضوعي/ الاختيار من متعدد، وعدم وجود توافق بين قدرة المفحوصين وصعوبة الفقرة وباختلاف كل من حجم العينة وطول الاختبار.

قام الباحث باستخدام برنامج (Wingen) لتوليد بيانات الدراسة وفق النموذج اللوجستي الثلاثي المعلمة، اما النماذج التي تم استخدامها في الدراسة فكانت: النموذج اللوجستي الثلاثي المعلمة ونموذج راش، وثلاثة نماذج معدّلة من نموذج راش الأول: نموذج راش بالإضافة لمعلمة تخمين ثابتة (c=0.20)، والثاني: نموذج راش مع نقطة القطع اللوغاريتمية Logarithm Cut (Cutlo=1) و (Cutlo=2) بهدف تعديل عدم التجانس بين معلمة الصعوبة ومعلمة القدرة في البيانات المولّدة.

وقام الباحث بتوليد البيانات لثلاث عينات مختلفة الحجم (n=200 , n=50 , n=1000) وبأطوال اختبارات مختلفة (k=40 , k=20) حيث n حجم العينة و k عدد فقرات الاختبار على التوالي، وجرى كذلك افتراض وجود حالتين للتخمين الأولى: تتوزع توزيعًا منتظماً بقيمة صغرى تساوي (٠,١٠) وقيمة عظمى تساوي (٠,٢٥) والحالة الثانية تفترض أنّ معلمة التخمين تساوي صفرًا. وبالنسبة لمعلمة القدرة فقد جرى توليدها وفق ثلاثة توزيعًات: الأولى تتوزع توزيعًا طبيعيًا بمتوسط (صفر) وانحراف معياري(١)، والثانية تتوزع توزيعًا طبيعيًا بمتوسط (-١) وانحراف معياري(١)، والحالة الثالثة تتوزع توزيعًا طبيعيًا بمتوسط (-٢) وانحراف معياري(١).

كما واستخدم الباحث برمجيتين هما: بارسكيل (Parascal) وون ستيب (Winstep) لتقدير معالم الفقرات، ومعلمة القدرة للمفحوصين وذلك باستخدام طريقة الأرجحية القصوى المشتركة. ولتقييم مدى دقة تقديرات المعالم(معلمة الصعوبة ومعلمة التمييز) ومعلمة القدرة للمفحوصين للنماذج الخمسة تم استخدام المحكات الآتية:

الأوساط الحسابية، والانحرافات المعيارية، والمئينات لمعلمة القدرة المولدة (Simulated Ability) وللقدرة المقدرة ولمعلمة الصعوبة، ومعامل ارتباط (كاندل تاو) بين قيم معلمة الصعوبة المولدة وتلك المقدرة، وبين قيم معلمة القدرة المقدّرة وتلك الموّلدة، والجذر التربيعي لمربعات الانحرافات

لمعلمة القدرة ومعلمة الصعوبة. وبينت نتائج الدراسة أنه في حالة النموذج اللوجستي الثلاثي المعلمة ونموذج تثبيت معلمة التخمين(c = 0) إذ كانت قيم الجذر التربيعي لمتوسط مربع الانحرافات لتقدير معلمة القدرة أقل منها للنماذج الاخرى ، وبلغت قيمة الجذر التربيعي لمتوسط مربع الانحرافات لتقدير معلمة القدرة (١,٩٩) في حالة عدم التوافق بين قيم معلمة القدرة ومعلمة الصعوبة، وتبين أنه بناءً على معاملات الارتباط بين الجذر التربيعي لمتوسط مربع الانحرافات ومتوسط مربع الانحرافات (RMSD) و (MSD) فإن النموذج اللوجستي الثلاثي المعلمة ونموذج تثبيت معلمة التخمين (c = 0) هما الخيار الأفضل لمعالجة التخمين في الحالات الآتية: حينما يكون هناك مواءمة جيدة بين كل من معلمة الصعوبة للفقرة ومعلمة القدرة للمفحوص، وبغض النظر عن وجود تخمين أو عدم وجوده، و حينما يكون حجم عينة المفحوصين قليلاً مع وجود تخمين.

كما قام جاو (Gao,2011) بدراسة هدفت لاستكشاف العلاقة بين التخمين وقدرة المفحوص من خلال كشف العلاقة بين التخمين الناجح الذي يقود إلى الإجابة الصحيحة والسمة الكامنة في إطار نماذج نظرية الاستجابة للفقرة؛ استخدم الباحث نموذجين جديدين من نماذج نظرية الاستجابة للفقرة هما: نموذج (راش للتخمين)، والنموذج (الثنائي للتخمين) اللذان يربطان بين احتمالية إجابة الفقرة إجابة صائبة مع معلمة القدرة للمفحوص ومعالم الفقرة.

حيث تمت مقارنة النموذج اللوجستي الثلاثي المعلمة مع النموذجين الجديدين من حيث دقة تقدير معالم الفقرات ومعلمة القدرة للمفحوصين، وجرى استخدام كل من أسلوب (مونتي كارلو) والأرجحية القصوى لتقدير معلمة القدرة للمفحوص، وتم الاعتماد على البيانات المولدة باستخدام الحزمة الاحصائية (SAS)، اما بالنسبة للمحكات التي استخدمت للمقارنة بين تقديرات المعالم فكانت: معامل الارتباط بين القيم المقدرة والقيم الحقيقية، والجذر التربيعي لمتوسط مربع الانحرافات المعيارية (RMSD) Root Mean Square of Standard Deviation والتحّيز (Bias).

إذ جرى توليد بيانات ثنائية تأخذ الإجابة عليها القيمة واحد في حالة الإجابة الصحيحة و (صفر) في حالة الإجابة الخطأ، ولعينات بحجم (١٠٠٠ مفحوص، و٥٠٠ مفحوص، و٢٠٠ مفحوص، و١٠٠ مفحوص) باستخدام النموذج اللوجستي الثلاثي المعلمة، والنموذج اللوجستي الثنائي المعلمة للتخمين، واستخدم ثلاثة اختبارات بأطوال مختلفة (٤٠ فقرة، و ٣٠ فقرة، و٢٠ فقرة) على التوالي حيث تم توليد معالم لهذه الفقرات ثم تثبيتها عبر أطوال الاختبارات الثلاثة، كما تم توليد معلمة القدرة بمتوسط (صفر) وانحراف معياري (١) وتم تكرار البيانات (١٠) مرات لتقليل خطأ المعاينة، أما بيانات المفحوصين فقد تم توليدها بناءً على كل من معالم الفقرة ومعلمة القدرة للمفحوصين السابقة.

أظهرت النتائج ارتباطًا عاليًا بين المعالم الحقيقية والمعالم المقدرة عند استخدام النموذج اللوجستي الثنائي للتخمين، كما كانت قيم الجذر التربيعي لمتوسط مربعات الانحرافات (RMSD) ومؤشر التحيز في دقة تقدير معالم الفقرات أقل منها في حالة النماذج الاخرى، كما كانت مطابقة هذا النموذج أفضل للبيانات، وكانت تقديرات معلمة القدرة ومعلمة التمييز أفضل في حالة النماذج الأخرى المستخدمة، اما النموذج اللوجستي الثلاثي المعلمة فقد اظهر تقديراً أفضل لمعلمة الصعوبة من النماذج الأخرى المستخدمة في دراسة الباحث.

استخدمت الدراسة بيانات مولدة باستخدام أسلوب المحاكاة من خلال برنامج نوهارم (Noharm)، حيث تم توليد بيانات بأربعة احجام مختلفة هي (١٠٠ مفحوص و ٢٥٠ مفحوصًا، و١٠٠٠مفحوص)، ثم قام الباحثون باستخدام برنامج بايلوج (Bilog MG) لتقدير كل من معالم الفقرات ومعلمة القدرة للمفحوصين للنماذج الستة المستخدمة المعدلة وغير المعدلة واظهرت نتائج الدراسة انه في العينات الصغيرة الحجم من المفحوصين فإن دقة التقدير للنماذج غير المعدلة أفضل من تلك المعدلة، وبزيادة حجم عينة المفحوصين تزداد الدقة في تقدير معالم الفقرات ومعلمة القدرة، كما أظهرت نتائج الدراسة ان النموذج اللوجستي الثلاثي المعلمة افضل من حيث مطابقة البيانات للنموذج ولدقة التقدير لمعالم الفقرات أو معلمة القدرة للمفحوصين.

وفي دراسة أجراها الحاج(٢٠١١) هدفت إلى معرفة أثر التحكم بمعلمة التخمين في نماذج استجابة الفقرة على تقدير معالم الفقرات ومعلمة القدرة للأفراد، باستخدام النموذج الأحادي المعلمة والنموذج ثنائي المعلمة والنموذج ثلاثي المعلمة، وجرى استخدام النماذج المعدلة أحادي المعلمة والنموذج ثنائي المعلمة بافتراض ثابت غير صفري لمعامل التخمين لكل منها بثلاثة مستويات وهي (٠,١٠ و ٠,٢٠ و ٠,٣٠) حيث قام الباحث بالمقارنة بين النماذج المعدلة والنماذج غير المعدلة من خلال مقارنة كل من معالم الفقرات ومعلمة القدرة للأفراد، تكوّن مجتمع الدراسة من طلبة الصف التاسع الأساسي في المدارس الحكومية التابعة لمديرية التربية والتعليم لمنطقة إربد الأولى للعام الدراسي(٢٠٠٩/٢٠١٠) أما عينة الدراسة فتكونت من (٧٤٧) طالباً وطالبة موزعين على (١٦) مدرسة تّم اختيارها بالطريقة العشوائية الطبقية تبعًا لمتغير الجنس، من(٨) مدارس للذكور و(٨) مدارس للإناث.

تكونت أداة الدراسة من اختبار تحصيلي في مبحث الرياضيات للصف التاسع الأساسي من نوع الاختبار الموضوعي/الاختيار من متعدد بأربعة بدائل للفقرة، وبعدد فقرات يبلغ (٣٨) فقرة، وبعد تطبيق الاختبار تم التأكد من مطابقة البيانات المتحققة من الاختبار لافتراضات نظرية الاستجابة للفقرة استخدم الباحث الرزمة الإحصائية للعلوم الاجتماعية SPSS بالإضافة للبرمجيات التالية:

(NOHARM) و (LDID) وحلّلت الاستجابات باستخدام النماذج التالية: ثلاثة نماذج غير معدّلة وستة نماذج معدّلة، وذلك باستخدام البرنامج الإحصائي (BILOG-MG3) والرزمة الإحصائية للعلوم الاجتماعية (SPSS) وتطبيق الاكسل (EXCEL) في إجراء التحليلات الاحصائية وعمل الرسوم المختلفة اللازمة.

أشارت نتائج الدراسة إلى أن النموذج اللوجستي ثلاثي المعلمة كان أكثر توافقًا مع البيانات المتحققة من الاختبار مقارنة بالنماذج أحادية المعلمة المعدّلة والنماذج ثنائية المعلمة المعدّلة، من حيث عدد الفقرات التي طابقت النموذج وعدد المفحوصين الذين طابقوا النموذج، ومن حيث دقة تقديرات معالم الفقرات ومعلمة قدرة المفحوصين حيث تبين أن قيم دالة معلومات الفقرة ودالة معلومات الاختبار والخطأ المعياري لتقدير القدرة كانت أفضل في حالة النموذج اللوجستي الثلاثي المعلمة من النماذج المعدلة الأخرى، كما أظهرت النتائج أن النموذج أحادي المعلمة المعدل بوضع معلمة التخمين كثابت عند (٠,٢٠)، والنموذج اللوجستي الثنائي المعلمة المعدل بوضع معلمة التخمين كثابت عند (٠,٢٠) أكثر توافقًا مع الاختبار من باقي النماذج أحادية المعلمة والنماذج ثنائية المعلمة وذلك من حيث عدد الفقرات التي طابقت النموذج وعدد المفحوصين الذين طابقوا النموذج، كذلك أظهرت النتائج أَن النموذج أحادي المعلمة غير المعدّل والنموذج ثنائي المعلمة غير المعدّل هما الأفضل من حيث دالة معلومات الفقرة ودالة معلومات الاختبار والخطأ المعياري في تقدير معلمة القدرة، مما يعني أن دقة التقديرات في النماذج غير المعدّلة أفضل من النماذج المعدّلة.

وفي دراسة أجراها الشواورة (٢٠١٣) هدفت إلى بيان أي من طريقتي الأرجحية العظمى وطريقة بييز تُعد أكثر دقة في تقدير معالم الفقرات عند اختلاف كل: من حجم العينة وعدد الفقرات والنموذج اللوجستي المستخدم، ولتحقيق هذا الهدف تمّ توليد بيانات ثنائية الإجابة، وفقًا للنموذج اللوجستي الأحادي المعلمة، والنموذج اللوجستي الثنائي المعلمة وبحجم عينة بلغ (٢٥٠ مفحوص و ٥٠٠ مفحوص و ١٠٠٠ مفحوص و ١٥٠٠ مفحوص)، وعدد فقرات الاختبار (١٠ فقرات و ٤٠فقرة)، بعد ذلك جرى تقدير معالم الفقرات تحت كل موقف من المواقف السابقة (حجم العينة وعدد الفقرات والنموذج اللوجستي المستخدم) من خلال طرق تقدير معالم الفقرات : الأرجحية العظمى في التقدير باستخدام برنامج (Bilog-MG3) ، وطريقة بييز في التقدير باستخدام برنامج (WinBugs v1.4) وتمّ تحلّيل قيم معالم الفقرات ومعلمة القدرة وفقاً لطريقتي التقدير الارجحية العظمى وطريقة بييز بواسطة مؤشرات دقة التقدير: (الجذر التربيعي لمتوسط مربعات الأخطاء، ومعامل الارتباط بين قيم المعالم الحقيقية والمعالم المقدرة) وفقاً لكل طريق من طرق التقدير المستخدمة: طريقة الأرجحية العظمى وطريقة بييز.

وبعد ذلك أجرى الباحث اختبار (T-Test) للمقارنة بين معاملات الارتباط لطريقتي التقدير، وتمّ إيجاد مؤشر الكفاءة النسبية (Relative Efficiency) ، وأظهرت نتائج الدراسة من خلال المحكات المستخدمة أفضلية لطريقة التقدير البييزية، وهذا ما أكدته نتائج اختبار (T-Test) حيث أظهرت معاملات الارتباط بين المعالم الحقيقية وتلك المقدرة من خلال طريقة بييز فروقًا دالة إحصائياً عند مستوى دلالة (α= (0.05 في معظم مواقف الدراسة، كما خلصت نتائج الدراسة إلى أن طريقة بييز (Bayes) للتقدير أظهرت دقة في تقدير معالم الفقرات مقارنة مع طريقة الأرجحية العظمى، وخاصة في العينات الصغيرة، أو عند استخدام النموذج اللوجستي الثنائي المعلمّة.

أجرى عبد الوهاب(٢٠١٣) دراسة هدفت للمقارنة بين النظرية الكلاسيكية في القياس والنموذج اللوجستي الثلاثي المعلمة في معالجة أثر التخمين عند تقدير معالم الفقرات ومعلمة القدرة على عينة من طلاب الفرقة الرابعة والدبلوم العام بكلية التربية بالمينا على اختبار تحصيلي من نوع الاختيار من متعدد في مادة القياس النفسي في أربعة مستويات معرفية، هي: الحفظ والفهم والتطبيق والتحليل؛ وذلك من أجل معرفة الفروق بين هذه الإحصائيات في ضوء النظرية الكلاسيكية في القياس قبل وبعد تصحيح أثر التخمين، وفي ضوء النموذج اللوجستي الثلاثي المعلمة في إطار نظرية الاستجابة للفقرة، استخدم الباحث برنامج (Param-3pl) لتقدير معالم الفقرات والافراد في النموذج اللوجستي الثلاثي المعلمة.

وللإجابة عن أسئلة الدراسة قارن الباحث بين معاملات الارتباط بين إحصائيات الفقرة قبل التصحيح من أثر التخمين في النظرية الكلاسيكية في القياس وبعد التصحيح، كما قارن بين معالم الفقرات والأفراد في ضوء النموذج اللوجستي الثلاثي المعلمة من خلال تحويل معاملات التمييز المقدرة في ضوء النظرية الكلاسيكية إلى قيمة (Z- Fisher) كما تم حساب معامل الارتباط بين إحصائيات الفقرات المقدرّة في ضوء النظرية الكلاسيكية في القياس قبل التصحيح من أثر التخمين وبعد التصحيح من أثر التخمين، وتم إيجاد معاملات الارتباط بين معلمة التمييز المقدرة من خلال النموذج اللوجستي الثلاثي المعلمة وتلك المستخرجة قبل التصحيح من أثر التخمين في ضوء النظرية الكلاسيكية في القياس، وأوضحت نتائج الدراسة عدم وجود اختلاف في تقدير إحصائيات الأفراد والفقرات في ضوء النظرية الكلاسيكية قبل وبعد تصحيح أثر التخمين، ولكن وجد اختلاف واضح في هذه الإحصائيات في ضوء النموذج الثلاثي المعلمة عن النظرية الكلاسيكية في القياس قبل التصحيح من أثر التخمين وبعده ممّا يُشير إلى اختلاف كل منهما في معالجة أثر التخمين.

كما قام الشريفين (٢٠١٢) بدراسة هدفت إلى الكشف عن أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة، ولتحقيق هدف الدراسة تم بناء اختبار تحصيلي في الفيزياء من نوع الاختيار من متعدد بأربعة بدائل تكون بصورته النهائية من (٣٣) فقرة، وطبق الاختبار على عينة الدراسة المكونة من (١٠٠٠) طالب وطالبة من طلبة الصف الثاني الثانوي العلمي، وحللت النتائج وفق النموذج الثلاثي المعلمة باستخدام البرمجية (Bilog-Mg)، وبينت النتائج وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) في متوسطات الأخطاء المعيارية لتقديرات معالم الفقرات تعزى للتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تعزى لمتغير حجم العينة أو طريقة التقدير.

كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) في متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تعزى لمتغير حجم العينة، وللتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تعزى لطريقة التقدير، كما بينت النتائج عدم وجود فروق دالة إحصائيا بين معاملات الثبات المقدرة وفق نظرية الاستجابة للفقرة عند أحجام العينة المختلفة (١٠٠ مفحوصًا و ٥٠٠ مفحوص، و ١٠٠٠ مفحوص)، وأشارت نتائج الدراسة إلى أن دقة تقديرات معلمة القدرة تزداد في حالة عينة الأفراد ذوي القدرة العالية، وعينة الأفراد ذوي القدرة المتدنية عند استخدام طريقة بييز التوقع والتعظيم (EAP)، في حين تزداد الدقة عند مستويات الأفراد ذوي القدرة المتوسطة باستخدام طريقة الأرجحية العظمى (MLE) بغض النظر عن حجم العينة.

التعقيب على الدراسات السابقة

ومن خلال الدراسات التي تم عرضها سابقًا، يُمكن استخلاص بعض الاتجاهات في النتائج التي أمكن التوصل إليها والمتعلقة بمعالجة التخمين ودقة تقدير معالم الفقرات ومعلمة القدرة، في الاختبارات الموضوعية/الاختيار من متعدد كما يلي:

- هناك قلة في الدراسات المحلّية التي تناولت التخمين في نظرية الاستجابة للفقرة، وقلة كذلك في الدراسات التي تناولت طريقة التقدير البييزية أيضًا، وبالتالي جاءت هذه الدراسة لتسلط الضوء على طريقة التقدير البييزية واستخدامها في تقدير معالم الفقرات ومعلمة القدرة للمفحوصين.

- إن دقة تقدير المعالم للفقرات والأفراد تتغير بشكل واضح تبعًا للطريقة المستخدمة في التقدير والبرنامج الحاسوبي أيضًا المستخدم في عملية التقدير؛ بعض الدراسات السابقة استخدمت أكثر من برنامج حاسوبي مع كل نموذج، ممّا يعني وجود اختلافات تُعزى لاختلاف البرامج الحاسوبية المُستخدمة لتقدير المعالم، بينما استخدمت الدراسة الحالية برنامجًا حاسوبيًا واحدًا هو برنامج (Winbug).

- لم تقم أي دراسة بفحص أثر تشكيلات مختلفة لحجم عينة المفحوصين وعدد الفقرات والنماذج اللوجستية المستخدمة، حيث أن النماذج التي تناولتها هذه الدراسة لم تتناولها أية دراسة أخرى مجتمعة.

- اكتفت الدراسات السابقة بالإحصاء الوصفي -خاصة الدراسات الأجنبية- للإجابة عن أسئلتها إذ استخدمت المتوسطات الحسابية والانحرافات المعيارية، أما هذه الدراسة فقد استخدمت الإحصاء الوصفي والاستدلالي مع ما يعُطيه من قوة للدراسة.

- وأخيرًا يمكن القول إن هذه الدراسة تتفق مع الدراسات السابقة في سعيها لتحري دقة تقدير معالم الفقرة ومعلمة القدرة وفي بعض النماذج المستخدمة، إلا أنها تجري مقارنة أوسع مما هو موجود في الدراسات السابقة، سواء من ناحية عدد النماذج المشمولة بالدراسة، أو تنوع ظروف التجربة من حيث عدد الفقرات وحجم العينة، وتختلف خصائص البيانات المولدة في هذه الدراسة عن غيرها من خصائص البيانات المولدة في الدراسات السابقة، كما تتميز بالتوسع في المحكات المستخدمة للحكم على فاعلية النماذج والإجراءات المستخدمة فيها.

الفصل الثالث

الطريقة والإجراءات

الفصل الثالث

الطريقة والإجراءات

يتضمن هذا الفصل وصفًا للإجراءات والبرامج التي استُخدمت لتوليد البيانات الخاصة بإيجاد القيم الحقيقية لمعالم الفقرة وقدرات المفحوصين، إذ يعرض وصفًا دقيقًا لطريقة توليد البيانات باستخدام برنامج ون جن (WinGen.3)، وكيفية تقدير معالم الفقرات (معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين)، ومعلمة القدرة للمفحوصين وفق طريقة بييز (Bayes) من خلال برنامج (WinBugs1.4)، حيث تم استخراج الإحصائيات التي تتطلبها الإجابة عن أسئلة الدراسة؛ والتي تتمحور حول فاعلية النماذج المستخدمة في هذه الدراسة في التغلب على مشكلة التخمين في الاختبارات الموضوعية/الاختيار من متعدد؛ وبالتالي التوصل الى تقديرات دقيقة لمعالم الفقرات ومعلمة القدرة للمفحوصين، وقد تمّ تقسيم إجراءات هذه الأطروحة إلى ثلاث مراحل رئيسة وفيما يلي وصفٌ لكل مرحلة:

أولاً: إجراءات توليد البيانات

تم توليد البيانات التي تمثل استجابات المفحوصين باستخدام برنامج ون جن(WinGen) وذلك وفقًا للمتغيرات الخاصة بالدراسة معالم الفقرات (معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين) ومعلمة القدرة للمفحوصين وطول الاختبار وعدد المفحوصين، إذ يتيح هذا البرنامج للمسُتخدم تحديد عدد الفقرات اللازمة وعدد الاستجابات ونوع النموذج اللوجستي المستخدم (أحادي المعلمة و ثنائي المعلمة وثلاثي المعلمة وغيرها من النماذج) مع إمكانية تخزين كل هذه البيانات في ملفات نصية لمعالجتها لاحقًا ونقلها للبرمجيات التي تستخدم لإجراء التحليل الاحصائي المناسب، كما يُتيح البرنامج إجراء تقدير لمعالم الفقرات ولمعلمة القدرة للمفحوصين للبيانات المولدة من خلال الربط مع برنامج (BilogMG3) أو برنامج (Parascale).

إذ جرى قبل توليد استجابات المفحوصين التحديد الآتي لخصائص كل من معالم الفقرات، ومعلمة القدرة للمفحوصين ليتم توليدها تبعًا لها، وذلك وفقًا للخطوات الآتية:

أ- توليد معلمة القدرة للمفحوصين بحيث تكون موزعة توزيعًا طبيعيًا بمتوسط حسابي مقداره (صفر) لوجت وانحراف معياري مقداره (١) لوجت.

ب- تحديد عدد الفقرات بحيث يتوافر للباحث اختباران: الأول (أ) بطول (٢٥) فقرة، والاختبار الثاني (ب) بطول (٥٠) فقرة، ويعود سبب اختيار الباحث لهذين الطولين إلى أن الاختبارات النفسية

والتربوية عادة ما تَستخدم اختبارات بطول (٢٥) فقرة كالاختبارات القصيرة وأخرى بطول (٥٠) فقرة كاختبارات متوسطة (Dawber,Roger and Carbonar,2009).

ج- توليد استجابات على كل اختبار من الاختبارين السابقين بحيث يكون عدد أفراد العينة الأولى (١٥٠٠) مفحوص وعدد أفراد العينة الثانية (٥٠٠) مفحوص ويُعزى اختيار هذين الحجمين المتفاوتين بهدف تحري اثر حجم العينة على معالجة التخمين وبالتالي تم اختيار عينة متوسطة الحجم، وعينة اخرى كبيرة الحجم وفي هذا السياق يُشير كل من هامبلتون وكوك (Hambleton and Cook, 1983) إلى أن حجم عينة يبلغ (٥٠٠) مفحوص يُعتبر ملائمًا لإعطاء تقديرات مستقرة لقدرات المفحوصين ومعالم الفقرات في النموذج اللوجستي الثلاثي المعلمة.

كما روعي عند اختيار خصائص معالم الفقرات قبل عملية توليد البيانات أن تكون مطابقة لافتراضات ومتطلبات النموذج اللوجستي الثلاثي المعلمة، بحيث تكون خصائص معالم الفقرات كما يلي:

١- معلمة التمييز: تتوزع توزيعًا منتظمًا بقيمة دنيا تساوي (٠,٤٠) وقيمة عليا تساوي (١,٧) لوجت.

٢- معلمة الصعوبة: تتوزع توزيعًا طبيعيًا بمتوسط حسابي (صفر) لوجت وانحراف معياري (١) لوجت.

٣- معلمة التخمين: تتوزع توزيعًا منتظمًا بقيمة دنيا تساوي (٠,٠٥) وقيمة عليا تساوي (٠,٤٠).

وبذلك تم توليد استجابات المفحوصين، وخُزنت في أربعة ملفات كما في التصميم الآتي:

جدول رقم(٢)

تصميم ملفات الاستجابات المولدة والمستخدمة في الدراسة

طول الاختبار / حجم العينة	الاختبار: اختبار(أ) (٢٥) فقرة	الاختبار: اختبار(ب) (٥٠) فقرة
٥٠٠ مفحوص	ملف١	ملف٢
١٥٠٠ مفحوص	ملف٣	ملف٤

تم تخزين كل من معالم الفقرات ومعلمة القدرة للمفحوصين واستجاباتهم على فقرات الاختبارات المولدة في ملفات نصية منفصلة، حيث تم استدعاؤها من خلال الرزمة الإحصائية للعلوم الاجتماعية (SPSS) وتم التعامل مع هذه الملفات على أنها تحوي المعالم الحقيقية للفقرات والمفحوصين True Items and Person Parameters.

ثانياً: التحقق من خصائص البيانات المولدة

بعد الانتهاء من عملية توليد البيانات، وترتيبها في ملفات نصية، حيث تم استخراج المتوسطات الحسابية لمعلمة القدرة ولمعالم الفقرات والانحرافات المعيارية المناظرة لها كما يُبينها الجدولان رقم(٣) ورقم (٤) على التوالي.

جدول (٣)

المتوسطات الحسابية لمعلمة القدرة للبيانات المولدة والانحرافات المعيارية المناظرة لها وفقًا لمتغيرات الدراسة (حجم العينة وطول الاختبار).

طول الاختبار / حجم العينة	الاختبار			
	اختبار (ا): (٢٥) فقرة		اختبار(ب) : (٥٠) فقرة	
	المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
٥٠٠ مفحوص	٠,٠١٣	١,٠٠٢	٠,٠٢٢	٠,٩٦
١٥٠٠ مفحوص	٠,٠١٥	٠,٩٧	٠,٠١١	١

يُلاحظ من الجدول (٣) أن توزيع معلمة القدرة للمفحوصين في الملفات الأربعة إلى حد ما توزيع سوي؛ وبشكل يضمن تحقق الشرط الذي تم وضعه لشكل توزيع معلمة القدرة في الحالات الأربع المختلفة الناتجة عن تقاطع متغيري أفراد العينة وفقرات الاختبار.

وللتأكد فيما إذا كانت الانحرافات البسيطة الملحوظة لقيم متوسطات معلمة القدرة وانحرافها المعياري تنحرف عن المتوسط الحسابي والانحراف المعياري للتوزيع السوي، تم استخدام الاختبار الاحصائي كولموغروف- سمرنوف (Kolmogorov- Smirnova) من خلال الرزمة الإحصائية للعلوم الاجتماعية؛ لفحص ذلك عند مستوى الدلالة (α=0.05)، (Arnold and Emerson ,2011) والجدول (٤) يُبين قيم النتائج المستخرجة.

جدول (٤)

نتائج اختبار كولمجروف-سمرنوف لفحص مدى تطابق المتوسط الحسابي لمعلمة القدرة المُولدة والانحراف المعياري المناظر لها مع المتوسط الحسابي للتوزيع السوي وانحرافه المعياري، تبعًا لمتغيري الاختبار وعدد المفحوصين.

ملف البيانات المولدة لمعلمة القدرة في حالة الاختبار.	قيمة الإحصائي (K-S)	مستوى الدلالة
٢٥ فقرة و ٥٠٠ مفحوص	٠,٠٢٦	٠,٢٠٠
٢٥ فقرة و ١٥٠٠ مفحوص	٠,٠٣٤	٠,٢٠١
٥٠ فقرة و ٥٠٠ مفحوص	٠,٠٢٩	٠,١٧٤
٥٠ فقرة و ١٥٠٠ مفحوص	٠,٠٣٦	٠,١٧٣

يُلاحظ من الجدول رقم(٤) أن جميع القيم المحسوبة للاختبار كولمجروف- سمرنوف، كان مستوى الدلالة لها يزيد عن (٠,٠٥)، الأمر الذي يُشير إلى أن معلمة القدرة المولدة للمفحوصين في حالة الملفات الأربعة الواردة في الجدول تتوزع توزيعًا سويًا، ممّا يؤكد تحقق التوزيع السوي لمعلمة القدرة في حالة هذه الملفات الأربعة.

وجرى كذلك استخراج المتوسطات الحسابية والانحرافات المعيارية المناظرة لمعالم الفقرات (معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين) بناءً على البيانات المولدة وذلك تبعًا لطول الاختبار وعدد المفحوصين، والجدول رقم(٥) يبين النتائج التي تم التوصل إليها.

جدول (٥)

المتوسطات الحسابية لمعالم الفقرات المولدة (الصعوبة، والتمييز، والتخمين) والانحرافات المعيارية المناظرة لها وفقًا لمتغيرات الدراسة حجم العينة وطول الاختبار.

طول الاختبار / حجم العينة		الاختبار					
		اختبار(أ) : ٢٥ فقرة			اختبار(أ) : ٥٠ فقرة		
		معلمة التمييز	معلمة الصعوبة	معلمة التخمين	معلمة التمييز	معلمة الصعوبة	معلمة التخمين
٥٠٠ مفحوص	متوسط حسابي	١,٠٢	٠,٠٢٤	٠,٢٥٠	١,٠٥٢	٠,٠١٩	٠,٢١٠
	انحراف معياري	٠,٤٠٤	٠,٨٢٦	٠,١٥١	٠,٣٥١	٠,٩٤٠	٠,١١٨
١٥٠٠ مفحوص	متوسط حسابي	١,٠٧٩	٠,٠١٧	٠,٢٨٤	١,١٣١	٠,٠٢٢	٠,٢٧٥
	انحراف معياري	٠,٤٣٤	٠,٨٦٦	٠,١٥٨	٠,٣٦٩	٠,٩١١	٠,١٥٠

يُلاحظ من الجدول (٥) أن المتوسطات الحسابية لمعلمة الصعوبة للفقرات في حالة كل من الاختبار المكون من (٢٥) فقرة والاختبار المكون من (٥٠) فقرة وفي حالة البيانات التي تُمثل (٥٠٠) مفحوص و(١٥٠٠) مفحوص قد انحرفت قليلاً عن المتوسط الحسابي للتوزيع السوي (الصفر) والانحراف المعياري له؛ فقد بلغ المتوسط الحسابي لمعلمة الصعوبة (٠,٠٢٤) لوجت وبانحراف معياري بلغ (٠,٨٢٦) لوجت في حالة الاختبار المكون من (٢٥) فقرة و (٥٠٠) مفحوص، اما في حالة الاختبار المكون من (٥٠) فقرة والمولد عليه استجابات لـ (٥٠٠) مفحوص فقد بلغ المتوسط الحسابي لمعلمة الصعوبة (٠,٠١٩) لوجت، وبانحراف معياري بلغ (٠,٩٤٠) لوجت، في حين كانت قيم المتوسط الحسابي لمعلمة الصعوبة في حالة الاختبار المكون من (٢٥) فقرة والمولد عليه استجابات لـ (١٥٠٠) مفحوص تساوي (٠,٠١٧) لوجت، وبانحراف معياري مقداره (٠,٨٦٦) لوجت، واخيرًا بلغ المتوسط الحسابي لمعلمة الصعوبة (٠,٠٢٢) لوجت، و

بانحراف معياري بلغ (٠,٩١١) لوجت، في حالة الاختبار المكون من (٥٠) فقرة والمولد عليه استجابات لـ (١٥٠٠) مفحوص.

بالنسبة لمعلمة التمييز والتي روعي عند توليد معالم الفقرات أن تتوزع توزيعًا منتظمًا فقد بلغ المتوسط الحسابي لها (١,٠٢) لوجت، وبانحراف معياري بلغ (٠,٤٠٤) لوجت في حالة الاختبار المكون من (٢٥) فقرة والمولد عليه استجابات لـ (٥٠٠) مفحوص، وبلغت قيم المتوسط الحسابي لمعلمة التمييز (١,٠٥٢) لوجت، وبانحراف معياري بلغ (٠,٣٥١) لوجت في حالة الاختبار المكون من (٥٠) فقرة والمولد عليه استجابات لـ (٥٠٠) مفحوص.

كما بلغت قيمة المتوسط الحسابي لمعلمة التمييز (١,٠٧٩) لوجت وبانحراف معياري بلغ(٠,٤٣٤) لوجت، في حالة الاختبار المكون من (٢٥) فقرة والمولد عليه استجابات لـ(١٥٠٠) مفحوص، وبلغت قيمة المتوسط الحسابي لمعلمة التمييز (١,١٣١) لوجت، وبانحراف معياري بلغ (٠,٣٦٩) لوجت في حالة الاختبار المكون من (٥٠) فقرة والمولد عليه استجابات لـ (١٥٠٠) مفحوص .

أما بالنسبة لمعلمة التخمين والتي روعي أن تتوزع توزيعًا منتظمًا عند توليد معالم الفقرات فقد بلغ المتوسط الحسابي لها(٠,٢٥٠) وبانحراف معياري بلغ (٠,١٥١) في حالة الاختبار المكون من (٢٥) فقرة والمولد عليه استجابات لـ (٥٠٠) مفحوص، بينما بلغ المتوسط الحسابي لمعلمة التخمين (٠,٢١٠) وبانحراف معياري بلغ (٠,١١٨)، في حالة الاختبار المكون من (٥٠) فقرة والمولد عليه استجابات لـ(٥٠٠) مفحوص، في حين بلغت قيمة المتوسط الحسابي لمعلمة التخمين (٠,٢٨٤) وبانحراف معياري بلغ (٠,١٥٨) في حالة الاختبار المكون من (٢٥) فقرة والمولد عليه استجابات لـ (١٥٠٠) مفحوص، واخيرًا بلغت قيمة المتوسط الحسابي لمعلمة التخمين (٠,٢٧٥) وبانحراف معياري مقداره (٠,١٥٠) في حالة الاختبار المكون من (٥٠) فقرة والمولد عليه استجابات لـ (١٥٠٠) مفحوص.

وللتأكد فيما إذا كانت الانحرافات في قيم المتوسطات الحسابية والانحرافات المعيارية المناظرة لها لمعلمة الصعوبة تطابق المتوسط الحسابي والانحراف المعياري للتوزيع السوي جرى استخراج نتائج اختبار كولمجروف- سمرنوف كما في الجدول رقم(٦).

جدول (٦)

نتائج كولمجروف- سمرنوف لفحص مدى تطابق المتوسطات الحسابية لتوزيع معلمة الصعوبة والانحرافات المعيارية المناظرة لها للبيانات المولدة مع المتوسط الحسابي للتوزيع السوي والانحراف المعياري له تبعًا لطول الاختبار وعدد المفحوصين.

ملف البيانات المولدة في حالة الاختبار	قيمة الإحصائي (K-S)	مستوى الدلالة
٢٥ فقرة و (٥٠٠) مفحوص	٠,١١٣	٠,١٩٧
٥٠ فقرة و(٥٠٠) مفحوص	٠,٠٨٧	٠,٢٠١
٢٥ فقرة و(١٥٠٠) مفحوص	٠,١١٥	٠,٢١٨
٥٠ فقرة و(١٥٠٠) مفحوص	٠,٠٩١	٠,٢٠٣

يُلاحظ من الجدول (٦) أن مستوى الدلالة الإحصائية لجميع قيم اختبار كولمجروف- سمرنوف في حالة ملفات البيانات المولدة الخاصة بمعلمة الصعوبة كانت تزيد عن (٠,٠٥) الأمر الذي يُشير إلى تطابق المتوسط الحسابي والانحراف المعياري لها مع المتوسط الحسابي للتوزيع السوي والانحراف المعياري له.

ثانيا : التحقق من افتراضات أُحادية البعد والاستقلال الموضعي:

تم التحقق من توافر شرط أحادية البعد كأحد افتراضات نظرية الاستجابة للفقرة للدرجات التي جرى توليدها للمفحوصين للاختبار (أ) المكون من (٢٥) فقرة والاختبار(ب) المكون من (٥٠) فقرة في حالة العينة المكونة من(٥٠٠) مفحوص و (١٥٠٠) مفحوص، باستخدام إجراءين هما: معامل الاتساق الداخلي/إحصائيات الفقرة باستخدام معادلة كرونباخ ألفا، وإجراء التحليل العاملي باستخدام طريقة المكونات الرئيسة، وفيما يلي عرضٌ للنتائج التي تم التوصل لها:

اولاً: معامل الثبات الاتساق الداخلي/ بدلالة إحصائيات الفقرة:

جرى استخراج معامل الثبات الاتساق الداخلي/إحصائيات الفقرة باستخدام معادلة كرونباخ ألفا للاختبار(أ) المكون من (٢٥) فقرة والاختبار (ب) المكون من (٥٠) فقرة باستخدام البيانات المولدة لكل منها في حالة العينة للبيانات المولدة بحجم (٥٠٠) مفحوص، وبحجم (١٥٠٠) مفحوص على التوالي: ويبين الجدول (٧) معاملات ثبات الاتساق الداخلي/إحصائيات الفقرة باستخدام معادلة كرونباخ ألفا لكل اختبار من هذين الاختبارين في كلا الحالتين:

الجدول رقم (٧)

معامل ثبات الاتساق الداخلي/ لإحصائيات الفقرة باستخدام معادلة كرونباخ ألفا للاختبارين من طول(٥٠) فقرة و (٢٥) فقرة في حالة البيانات المكونة من (٥٠٠) مفحوص و(١٥٠٠) مفحوص.

طول الاختبار / حجم العينة	الاختبار	
	الاختبار(ا): ٢٥ فقرة	الاختبار (ب) : ٥٠ فقرة
٥٠٠ مفحوص	٠,٨٦	٠,٨٨
١٥٠٠ مفحوص	٠,٩٠	٠,٩٢

يُلاحظ من الجدول (٧) أنّ جميع قيم معاملات الثبات الاتساق الداخلي/إحصائيات الفقرة باستخدام معادلة كرونباخ ألفا كانت فوق (٠,٨٦)، إذ يُشير كل من دينيك وتفكول (Dennick and Tavakol, 2011) إلى أن قيم معامل الثبات التي تزيد عن (٠,٧٠) تُعتبر مقبولة

لأغراض تحقق افتراض توفر أحادية البعد في الأداء على الاختبار، وأن قيم معامل الثبات التي تقل عن ذلك تُشير إلى ارتباطات ضعيفة بين فقرات الاختبار، ممّا يدل على عدم تجانسها.

ثانياً: التحقق من أحادية البعد من خلال إجراء التحليل العاملي

للتحقق من توافر أحادية البعد في الأداء على الاختبارات التي تم تطويرها، جرى استخدام الرزمة الإحصائية للعلوم الاجتماعية (SPSS)؛ لاستخراج نتائج التحليل العاملي باستخدام طريقة المُكونات الرئيسة لاستجابات المفحوصين، على فقرات الاختبارات المكونة من (٢٥) فقرة و (٥٠) فقرة في حالة عدد الاستجابات المولدة عليها (٥٠٠) مفحوص و (١٥٠٠) مفحوص، حيث جرى استخراج قيم الجذور الكامنة للعوامل المتحققة عليها، بالإضافة إلى نسب التباين المُفسّر لكل عامل من العوامل، وجرى التدوير باستخدام طريقة التدوير المتعامد (Varimax- Rotation) للعوامل التي كانت قيمة الجذر الكامن لها أكبر من الواحد.

وتبين الجداول (٨-١١) نتائج التحليل العاملي للإجابات المولدة على الاختبار المكون من (٥٠) فقرة و(٢٥) فقرة في حالة العينة التي المكونة من (٥٠٠) مفحوص والعينة المكونة من (١٥٠٠) مفحوص على التوالي.

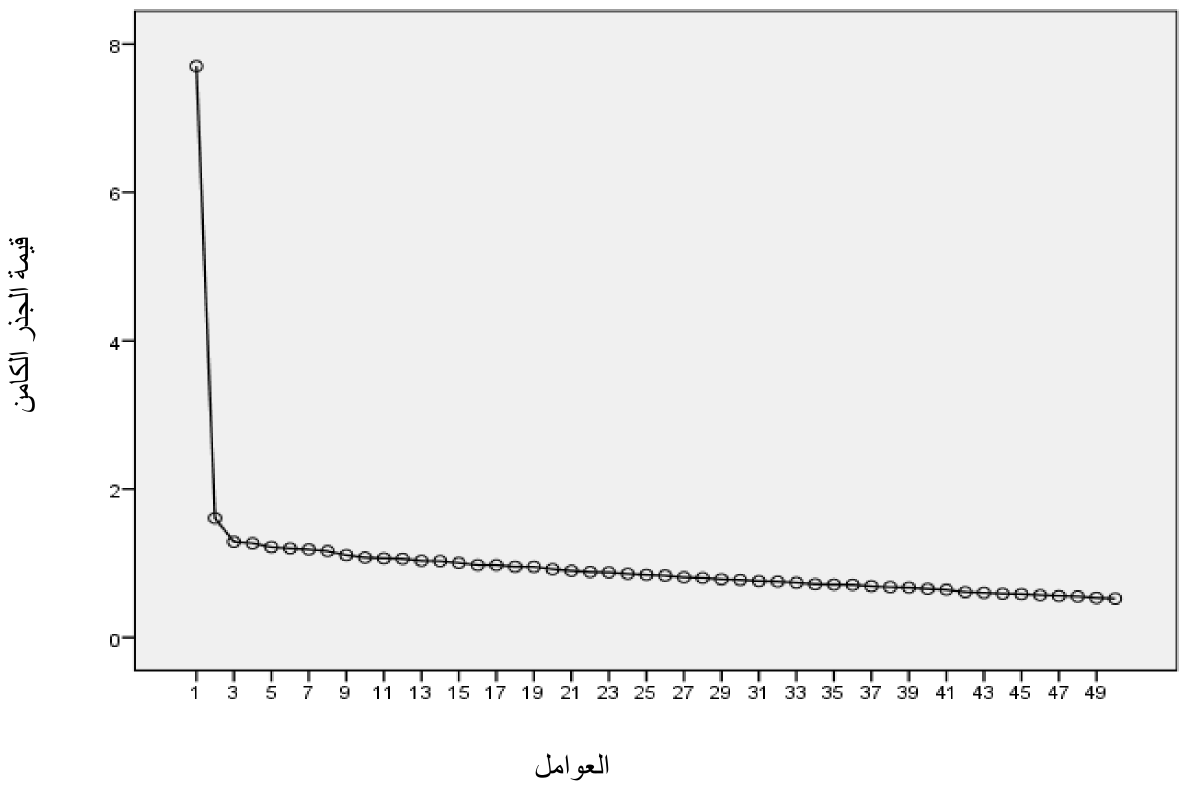
جدول رقم (٨)

ملخص نتائج التحليل العاملي باستخدام طريقة المكونات الرئيسة والتدوير المتعامد للأداء على الاختبار المكون من (٥٠) فقرة عندما تم توليد بيانات لعينة مكونة من (٥٠٠) مفحوص.

العامل	الجذر الكامن	نسبة التباين المفسرة	نسبة التباين التراكمية
١	٥,٩٦٠	١١,٩١٩	١١,٩١٩
٢	١,٥٤٤	٣,٠٨٩	١٥,٠٠٨
٣	١,٤٥٩	٢,٩١٨	١٧,٩٢٦
٤	١,٤١٠	٢,٨١٩	٢٠,٧٤٥
٥	١,٣٨٨	٢,٧٧٥	٢٣,٥٢١
٦	١,٣٢٩	٢,٦٥٧	٢٦,١٧٨

يُلاحظ من الجدول رقم(٨) أن هنالك ستة عوامل كانت الجذور الكامنة لها تزيد عن الواحد الصحيح؛ إذ أن العامل الأول فسّر ما نسبته (١١,٩١٩%) من التباين للأداء على الاختبار المكون من (٥٠) فقرة وهذه النسبة زادت عن (١١%) من التباين الكلي، كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن الثاني للعامل الثاني بلغت (٣,٨٦٠) ضعف، وهي نسبة كبيرة تزيد عن ضعفين الأمرالذي يُعّد مؤشرًا على تحقق سمة أحادية البعد في الأداء على هذا الاختبار في حالة الاختبار المكون من (٥٠) فقرة والإجابات المولدة عليه لـ (٥٠٠) مفحوص.

كما و يُبيّن الشكل رقم (١) تمثيلاً بيانيًا لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي للاختبار الذي طوله (٥٠) فقرة عندما جرى توليد البيانات عليه لعينة مكونة من (٥٠٠) مفحوص.



شكل رقم (١)

التمثيل البياني للجذور الكامنة للعوامل التي تفسر الأداء على الاختبار المكون من (٥٠) فقرة عندما جرى توليد بيانات عليه لعينة مكونة من (٥٠٠) مفحوص.

يُلاحظ من الشكل (١) أن مقدار الميل بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالميل بين العامل الثاني والعامل الثالث، ثم بقاء الميل بعد ذلك متقاربًا بين العوامل الباقية، ممّا يُشير إلى تحقق سمة أحادية البعد في تفسير الأداء على الاختبار.

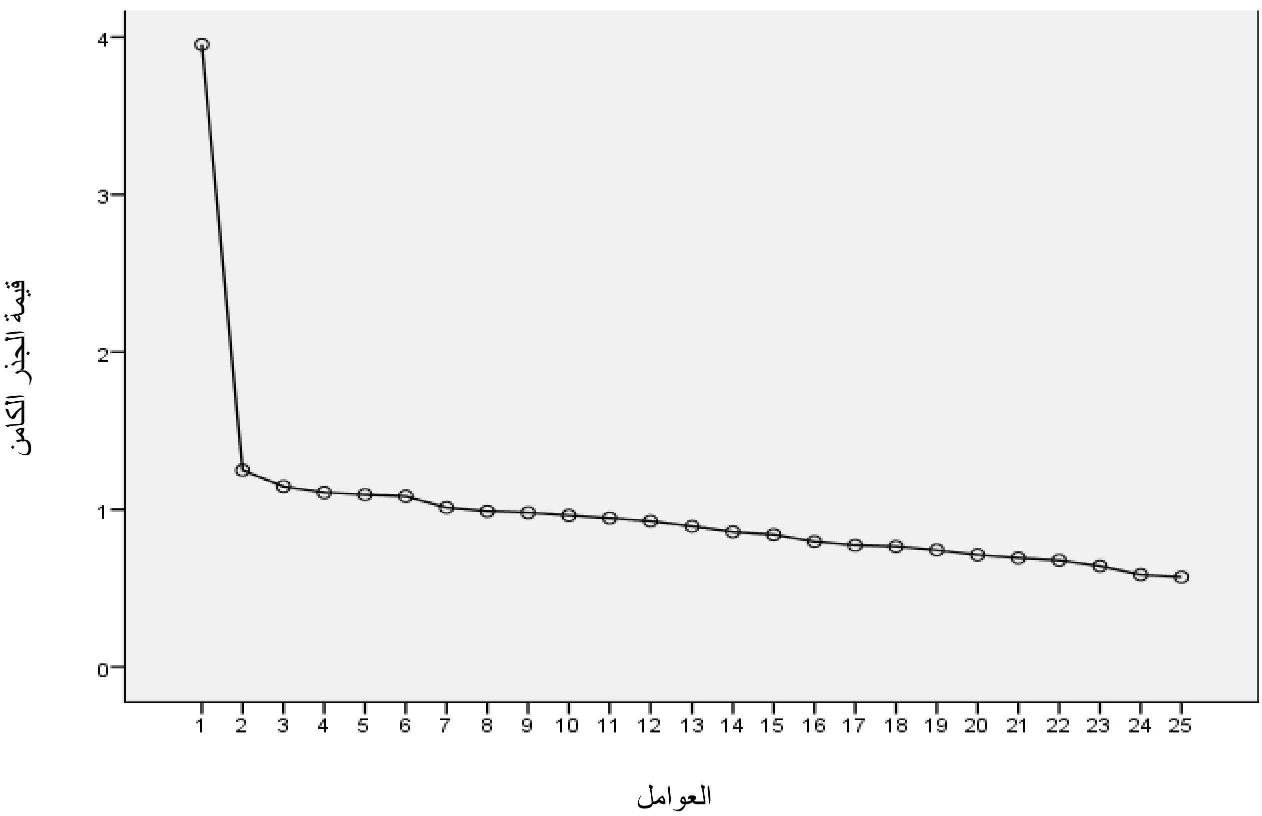
جدول رقم (٩)

ملخص نتائج التحليل العاملي باستخدام طريقة المكونات الرئيسة والتدوير المتعامد للأداء على الاختبار المكون من (٢٥) فقرة عندما تم توليد بيانات عليه لعينة مكونة من (٥٠٠) مفحوص.

العامل	الجذر الكامن	نسبة التباين المفسرة	نسبة التباين التراكمية
١	٣,٢٩١	١٣,١٦٥	١٣,١٦٥
٢	١,٣٠٧	٥,٢٢٧	١٨,٣٩٢
٣	١,٢٢٥	٤,٩٠٠	٢٣,٢٩٢
٤	١,١٩٢	٤,٧٧٠	٢٨,٠٦٢
٥	١,١٣٦	٤,٥٤٣	٣٢,٦٠٥
٦	١,٠٧٣	٤,٢٩٣	٣٦,٨٩٨
٧	١,٠٥٦	٤,٢٢٦	٤١,١٢٤
٨	١,٠٣٥	٤,١٤٠	٤٥,٢٦٤

يُلاحظ من الجدول رقم(٩) أن هناك ثمانية عوامل كانت الجذور الكامنة لها تزيد عن الواحد الصحيح وأن العامل الأول فسر ما نسبته (١٣,١٦٥%) من التباين للأداء على الاختبار المكون من (٢٥) فقرة، وهذه النسبة زادت عن (١٣%) من التباين، كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت (٢,٥١٨) ضعف، وهي نسبة تزيد عن ضعفين، وهذا يعتبر مؤشرًا على تحقق أحادية البعد للبيانات في الاختبار المكون من (٢٥) فقرة والمُولد عليه استجابات لـ(٥٠٠) مفحوص.

كما ويبين الشكل رقم (٢) التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي للاختبار الذي طوله (٢٥) فقرة عندما جرى توليد البيانات عليه لعينة مكونة من (٥٠٠) مفحوص.



شكل رقم (٢)

التمثيل البياني للجذور الكامنة للعوامل التي تفسر الأداء على الاختبار المكون من (٢٥) فقرة عندما جرى توليد بيانات عليه لعينة مكوّنة من (٥٠٠) مفحوص.

يُلاحظ من الشكل (٢) أن مقدار الميل بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالميل بين العامل الثاني والعامل الثالث، ثم بقاء الميل بعد ذلك متقاربًا بين العوامل الباقية، ممّا يُشير إلى تحقق افتراض أحادية البعد في تفسير الأداء على الاختبار للإجابات المولدة لـ (٥٠٠) مفحوص.

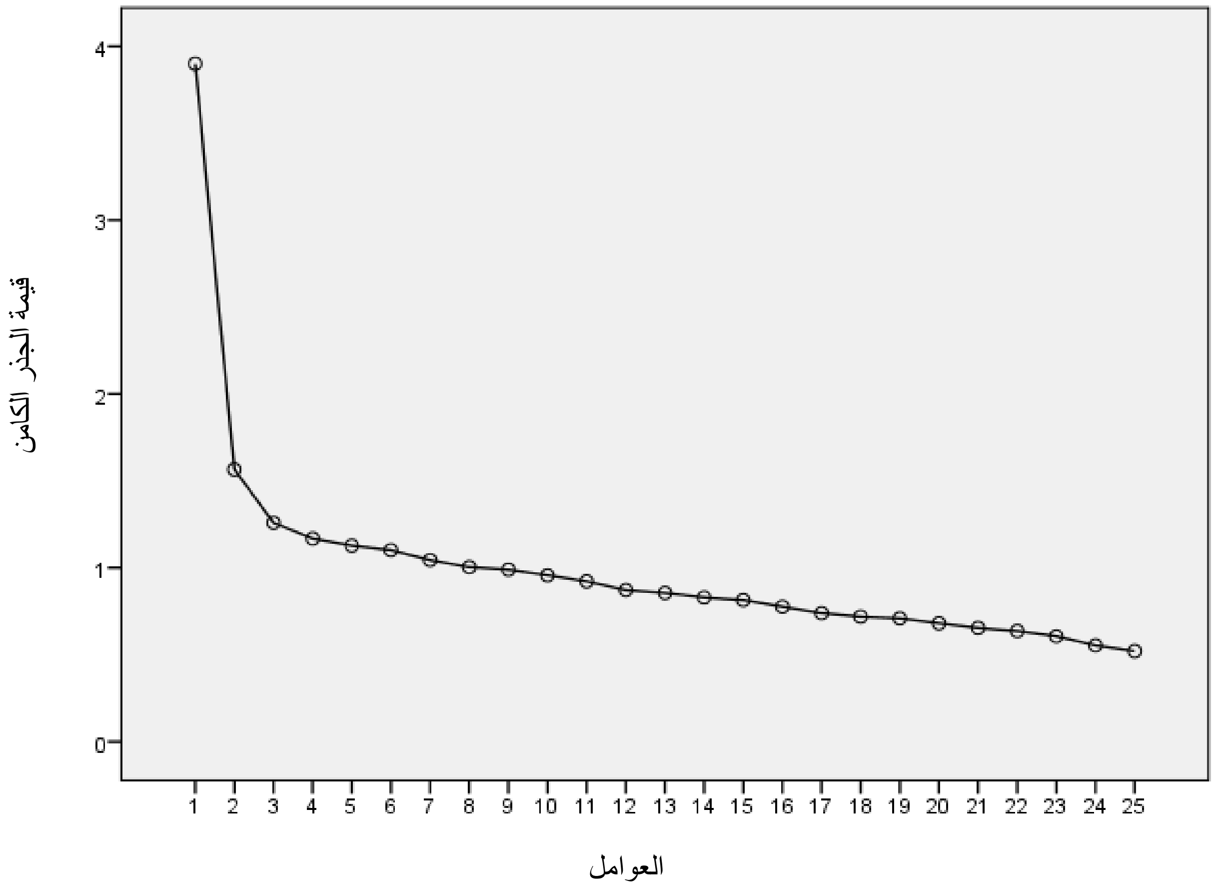
جدول رقم (١٠)

ملخص نتائج التحليل العاملي باستخدام طريقة المكونات الرئيسة والتدوير المتعامد للأداء على الاختبار المكون من (٢٥) فقرة عندما تم توليد بيانات عليه لعينة مكونة من (١٥٠٠) مفحوص.

العامل	الجذر الكامن	نسبة التباين المفسرة	نسبة التباين التراكمية
١	٣,٤٥٤	١٣,٨١٧	١٣,٨١٧
٢	١,١٦٠	٤,٦٣٨	١٨,٤٥٦
٣	١,١٢٢	٤,٤٨٨	٢٢,٩٤٤
٤	١,٠٨٩	٤,٣٥٧	٢٧,٣٠٠
٥	١,٠٦٥	٤,٢٦٢	٣١,٥٦٢
٦	١,٠٥١	٤,٢٠٣	٣٥,٧٦٥
٧	١,٠٣٥	٤,١٣٨	٣٩,٩٠٣
٨	١,٠١٥	٤,٠٥٩	٤٣,٩٦٢

يُلاحظ من الجدول رقم(١٠) أن هناك ثمانية عوامل يزيد الجذر الكامن لها عن واحد، وأن العامل الأول فسر ما نسبته (١٣,٨١٧%) من التباين للأداء على الاختبار المكون من (٢٥) فقرة وهذه النسبة زادت عن (١٣%) من التباين، كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت (٢,٩٧٧) ضِعف، وهي نسبة تزيد عن ضعفين، وهذا يُعتبر مؤشرًا على تحقق أحادية البعد للبيانات في الاختبار في حالة الملف الذي يحوي (١٥٠٠) مفحوص وعدد فقرات (٢٥) فقرة.

كما ويبين الشكل رقم (٣) التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي المتحققة للاختبار الذي طوله (٢٥) فقرة عندما جرى توليد البيانات عليه باستخدام عينة مكونة من (١٥٠٠) مفحوص.



شكل رقم (٣)

التمثيل البياني للجذور الكامنة للعوامل التي تفسّر الأداء على الاختبار المكون من (٢٥) فقرة عندما جرى توليد بيانات عليه لعينة مكونة من (١٥٠٠) مفحوص.

يُلاحظ من الشكل (٣) أن مقدار الميل بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالميل بين العامل الثاني والعامل الثالث، ثم بقاء الميل بعد ذلك متقاربًا بين العوامل الباقية، ممّا يشير إلى تحقق افتراض أحادية البعد في تفسير الأداء على الاختبار للإجابات المولدة لـ(١٥٠٠) مفحوص.

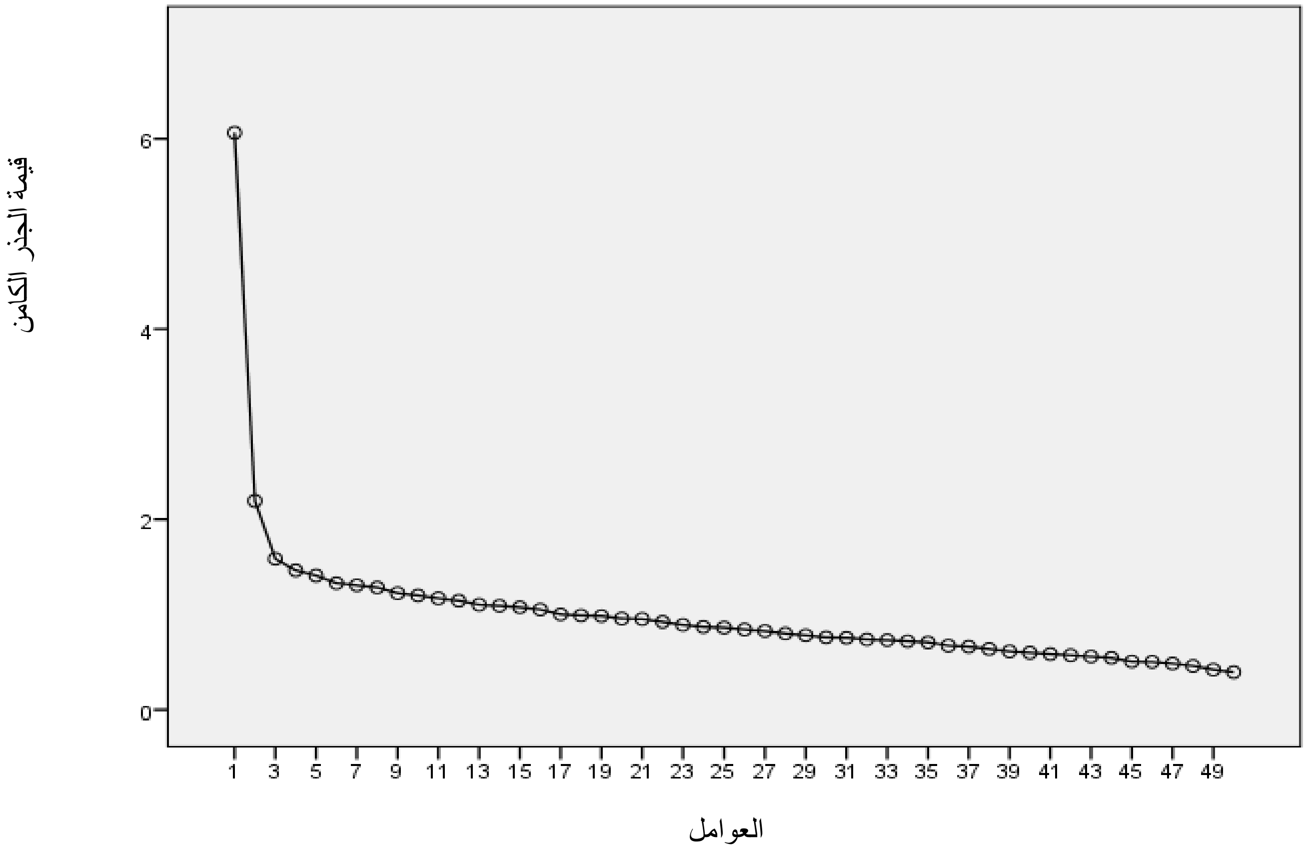
الجدول رقم (١١)

ملخص نتائج التحليل العاملي باستخدام طريقة المكونات الرئيسة والتدوير المتعامد للأداء على الاختبار المكون من (٥٠) فقرة عندما تم توليد بيانات عليه لعينة مكونة من (١٥٠٠) مفحوص.

العامل	الجذر الكامن	نسبة التباين المفسرة	نسبة التباين التراكمية
١	٥,٨٩٠	١١,٧٧٩	١١,٧٧٩
٢	١,٢٧٤	٢,٥٤٩	١٤,٣٢٨
٣	١,٢٣١	٢,٤٦٣	١٦,٧٩٠
٤	١,٢٠٠	٢,٤٠٠	١٩,١٩٠
٥	١,١٨٦	٢,٣٧٢	٢١,٥٦٢
٦	١,١٧٢	٢,٣٤٣	٢٣,٩٠٥
٧	١,١٣٤	٢,٢٦٨	٢٦,١٧٣
٨	١,١٢٦	٢,٢٥٣	٢٨,٤٢٦

يُلاحظ من الجدول رقم(١١) أن هناك ثمانية عوامل يزيد الجذر الكامن لها عن الواحد الصحيح وأن العامل الأول فسّر ما نسبته (١١,٧٧٩%) من التباين للأداء على الاختبار المكون من (٥٠) فقرة وهذه النسبة زادت عن (١١%) من التباين، كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت (٤,٦٢٣) ضعف، وهي نسبة كبيرة تزيد عن ضعفين وهذا يُعتبر مؤشرًا على تحقق أحادية البعد للبيانات في هذا الاختبار في حالة الملف الذي يحوي (١٥٠٠) مفحوص وعدد فقرات (٥٠) فقرة.

كما ويُبين الشكل رقم(٤) التمثيل البياني لقيم الجذور الكامنة للعوامل الناتجة عن التحليل العاملي للاختبار الذي طوله (٥٠) فقرة عندما جرى توليد البيانات عليه باستخدام عينة مكونة من (١٥٠٠) مفحوص.



شكل رقم(٤)

التمثيل البياني للجذور الكامنة للعوامل التي تفسر الأداء على الاختبار المكون من (٥٠) فقرة عندما جرى توليد بيانات عليه لعينة مكونة من (١٥٠٠) مفحوص.

يُلاحظ من الشكل (٤) أن مقدار الميل بين العامل الأول والعامل الثاني كان كبيرًا جدًا إذا ما قورن بالميل بين العامل الثاني والعامل الثالث، ثم بقاء الميل بعد ذلك متقاربًا بين العوامل الباقية، ممّا يُشير إلى تحقق افتراض أحادية البعد في تفسير الأداء على الاختبار للإجابات المولدة لـ (١٥٠٠) مفحوص.

ونظرًا لتحقق افتراض أحادية البعد في البيانات المولدة فهذا يعني تحقق افتراض الاستقلال الموضــعي Local Indpendence الذي يُعد مؤشــرًا لأحادية البعد إذا كان النموذج المســتخدم يحدد قيمة تقديرية لقدرة الفرد على بعد أحادي (Hamblton) and Swaminathan, 1985).

ثالثا: تقدير معالم الفقرة ومعلمة القدرة للمفحوصين (المعالم المقدرة):

تم في هذه المرحلة استدعاء استجابات المفحوصين الممثلة بالملفات الأربعة التي تم توليدها باستخدام برنامج ون جن، وتم استخدم برنامج (Winbug) بهدف تقدير معالم الفقرات ومعلمة القدرة للمفحوصين وفق كل نموذج من النماذج التي جرى توظيفها في إطار هذه الأطروحة لمعالجة التخمين وهي: النموذج اللوجستي الثلاثي المعلمة، والنموذج الثنائي المعدل إلى ثلاثي، ونموذج عتبة التخمين والنموذج القائم على الصعوبة، وباختلاف كلّ من عدد فقرات الاختبار(٢٥ فقرة و٥٠ فقرة) وحجم العينة (٥٠٠ مفحوص و١٥٠٠مفحوص).

مرت عملية تقدير البيانات المولدة في ثلاث مراحل رئيسة هي: المرحلة التكيفية (Adaptive Stage)، إذ يتم فيها تقدير معالم الفقرات ومعلمة القدرة تقديرًا أوليًا، يليها مرحلة الحرق (Burn In Stage) وفيها يتم استبعاد التقديرات الناتجة في المرحلة السابقة ثم البدء من جديد من خلال مرحلة التقدير النهائية (Final Estimating Stage) والتي يتم التوصل بعد انتهائها الى التقديرات النهائية لكل من معالم الفقرات ومعلمة القدرة للمفحوصين، حيث يتم عرض كل منها في ملفات نصية. نتج عن هذه المرحلة (١٦) ملفًا للقدرة ولمعالم الفقرات، يحتوي كل واحد منها على تقديرات لمعالم الفقرات ولمعلمة القدرة للمفحوصين، وتم تخزينها في ملفات منفصلة واستخدمت هذه البيانات في عملية التحليل للإجابة عن أسئلة الدراسة باستخدام الرزمة الإحصائية للعلوم الاجتماعية، والملحق رقم (٢) يُبين خطوات عملية التقدير هذه باستخدام برمجية (Winbug).

الإجراءات الإحصائية

أستخدمت الإجراءات الإحصائية التالية للإجابة عن كل سؤال من أسئلة الدراسة بعد ان تم الانتهاء من تقدير معالم الفقرات وقدرات المفحوصين:

السؤال الأول والثاني: للإجابة عن سؤال الدراسة الأول والثاني واللذين يتناولان فيما إذا كانت هنالك اختلافات في دقة تقدير معلمة القدرة ومعالم الفقرات(الصعوبة والتمييز والتخمين) على التوالي باختلاف النماذج المستخدمة لمعالجة التخمين في إطار هذه الدراسة وتبعًا لمتغير حجم العينة وعدد فقرات الاختبار جرى استخدام محكين للدقة: الأول الجذر التربيعي لمتوسط مربعات الخطأ والثاني مؤشر التحيز، حيث جرى استخراج المتوسطات الحسابية والانحرافات المعيارية لهما في حالة كل معلم من معالم الفقرات (الصعوبة، والتمييز، والتخمين) ومعلمة القدرة وجرى فحص الاختلافات الملحوظة بين متوسطات الجذر التربيعي لمتوسط مربعات الخطأ ومؤشر التحيز تبعًا لنموذج معالجة التخمين وطول الاختبار وحجم العينة باستخدام تحليل التباين الثلاثي للقياسات المتكررة متبعًا باختبار

اقل فرق دال (LCD) Least Significant Difference للمقارنات البعدية للوقوف على دلالة الفروق للمقارنات الثنائية في حالة القيمة الدالة لـ (ف) في حالة متغير نموذج معالجة التخمين.

للإجابة عن سؤال الدراسة الثالث:

تم حساب التباين للقيم المقدرة وفقاً لكل نموذج من النماذج المستخدمة في هذه الدراسة لمعالجة التخمين في المواقف المختلفة للمتغيرات (طول الاختبار وحجم عينة البيانات المولدة) على تباين النماذج الأخرى لكل معلم من معالم الفقرات ولمعلمة القدرة للمفحوصين، وفق المعادلة التالية:

$$RE(\theta) = \frac{I_a(\theta)}{I_b(\theta)} = \left(\frac{SE_b(\theta)}{SE_a(\theta)}\right)^2$$

فإذا زادت هذه القيمة عن الواحد الصحيح تكون الأفضلية للاختبار او النموذج (a) وإذا قلت عن الواحد تكون الأفضلية للاختبار او النموذج (b).

الفصل الرابع

نتائج الدراسة

الفصل الرابع

عرض نتائج الدراسة

هدفت هذه الدراســـة إلى التحقق من مدى فاعلية أربعة نماذج لمعالجة التخمين تندرج في إطار النموذج اللوجســـتي الثلاثي المعلمة: النموذج الثنائي المعلمة المعدل والنموذج اللوجســـتي الثلاثي المعلمة، ونموذج عتبة التخمين، والنموذج القائم على الصعوبة، بهدف التعرف على مدى فـاعليـة كـل نموذج من هـذه النمـاذج في معـالجـة التخمين المتضـــمن في فقرات الاختبـارات الموضـــوعية/الاختيار من متعدد عن طريق تحديد ايها يترك آثارًا ســـلبية أقل على تقدير كل من معلمة القدرة ومعالم الفقرات تحت ظروفٍ مختلفة كطول الاختبار، وحجم عينة المفحوصـــين، إذ تمّ استخدام برنامج (WinGen3) لتوليد استجابات للمفحوصين على اختبارين، الأول تضمّن (٢٥) فقرة والثـاني (٥٠) فقرة، تمّ توليـدهـا بحيث يُراعى أن تكون معـالم الفقرات ومعلمـة القـدرة لهـا خصـــائص محددة وقد تم فحص توفرها في إطار الفصـــل الثالث من هذه الأطروحة إذ تم توليد بيـانـات لعينتين على كـل اختبـار من الاختبـارين الســـابقين بواقع (٥٠٠) مفحوص و(١٥٠٠) مفحوص على التوالي.

وفيما يلي عرضٌ لأبرز النتائج حيث وضعت في جزئيين هما: النتائج العامة والنتائج التي تجيب على أسئلة الدراسة.

النتائج العامة المتعلقة بتقدير معالم الفقرات(الصـــعوبة، والتميز، والتخمين) بإختلاف متغيرات الدراسة (النموذج المستخدم لمعالجة التخمين، وعدد الفقرات، وحجم العينة).

للوقوف على مدى الاختلافات في تقدير معالم الفقرات ومعلمة القدرة باختلاف النموذج المســـتخدم في معالجة معلمة التخمين (النموذج اللوجستي الثلاثي المعلمة، والنموذج اللوجستي الثنائي المعلمة المعدل، والنموذج القائم على الصـــعوبة، ونموذج عتبة التخمين) تبعًا لطول الاختبار (٢٥ فقرة و٥٠ فقرة) وحجم العينة المستخدم (٥٠٠ مفحوص و ١٥٠٠ مفحوص)، تمّ حساب المتوسطات الحســـابية والانحرافات المعيارية المناظرة لها لمعالم الفقرات (معلمة الصـــعوبة ومعلمة التمييز ومعلمة التخمين) ومعلمة القدرة للمفحوصين التي تبدو في الجدول (١٢).

جدول رقم (١٢)

المتوسطات الحسابية والانحرافات المعيارية المناظرة لها لمعالم الفقرات (المقدرة) باستخدام البيانات المولدة بحجم (٥٠٠) مفحوص و (١٥٠٠) مفحوص على الاختبارين بحجم (٢٥) فقرة و(٥٠) فقرة وباختلاف أربعة نماذج لمعالجة التخمين.

حجم العينة	النموذج	طول الاختبار اختبار (أ) ٢٥ فقرة معلمة التمييز	معلمة الصعوبة	معلمة التخمين	معلمة القدرة	اختبار (ب) ٥٠ فقرة معلمة التمييز	معلمة الصعوبة	معلمة التخمين	معلمة القدرة
٥٠٠ مفحوص	الثلاثي	م=١,٠٥٢ ع=٠,٤٤٤	م=٠,٠٣٠ ع=٠,٦١٢	م=٠,١٩٢ ع=٠,٠٧٦	م=٠,٠١٣ ع=٠,٨٥٣	م=١,٠١٧ ع=٠,٤١٩	م=٠,١٠٥ ع=٠,٨٢٦	م=٠,١٧٣ ع=٠,٠٨٥	م=٠,٠٢٣ ع=٠,٩٢٣
	الثنائي المعدل	م=١,٠٨٨ ع=٠,٥١٢	م=٠,٠٦٩ - ع=٠,٧٧٧	*	م=٠,٠١٠ ع=٠,٩٠٠	م=١,٠٥١ ع=٠,٥٥٤	م=٠,٠٧٢ ع=٠,٩٣٨	*	م=٠,٠١٠ ع=٠,٩١٤
	عتبة التخمين	م=٠,٧٨٦ ع=٠,٣٦٩	م=-٠,٧١٣ ع=٠,٨٠٥	م=٠,٢١٤ ع=٠,١٤٦	م=٠,٠٣٧ ع=٠,٨١٩	م=٠,٨٠٧ ع=٠,٣٦٦	م=-٠,٦٢٥ ع=٠,٩٢٢	م=٠,٢٢١ ع=٠,١٢٢	م=٠,٠٦٣ ع=٠,٨٩١
	القائم على الصعوبة	م=٠,٩٥٩ ع=٠,٤٢٢	م=-٠,٤٢٧ ع=٠,٦١١	م=٠,٢٤٥ ع=٠,١٤٧	م=٠,٠٢٦ ع=٠,٨٣٤	م=٠,٨٥٧ ع=٠,٤٠٣	م=-٠,٦٢٣ ع=٠,٨٦١	م=٠,٢١٤ ع=٠٠,١٢٠	م=٠,٠٧ ع=٠,٨٨٥
١٥٠٠ مفحوص	الثلاثي	م=٠,٩٩١ ع=٠,٤٠٥	م=-٠,٢٧٥ ع=٠,٦٧٢	م=٠,٢٣٩ ع=٠,١٦٤	م=-٠,٠٠٣ ع=٠,٨٦٧	م=١,٠٤٠ ع=٠,٣٦٧	م=-٠,٢٨٩ ع=٠,٨٤٠	م=٠,٢١٣ ع=٠,١١٣	م=٠,٠١٣ ع=٠,٩٢١
	الثنائي المعدل	م=١,٠٠١ ع=٠,٥٢٢	م=-٠,٥٦٠ ع=١,٠٢١	*	م=٠,٠١١ ع=٠,٩٠٠	م=١,٠١٥ ع=٠,٤٩٧	م=-٠,٤٠٥ ع=٠,٩٧٨	*	م=٠,٠١٢ ع=٠,٩١٨
	عتبة التخمين	م=٠,٨٠١ ع=٠,٣٣٨	م=-١,١٧٦ ع=١,٠٣١	م=٠,٢٢٤ ع=٠,١٢٧	م=٠,٠١٩ ع= ٠,٨٥٢	م=٠,٨١٢ ع=٠,٣٠٩	م=-٠,٩٦٩ ع=٠,٩٦٦	م=٠,٢٧٧ ع=٠,٢٧٨	م=٠,٠٣٢ ع=٠,٩٠٢
	القائم على الصعوبة	م=١,١٠٢ ع=٠,٤٤٤	م=-٠,٦٨٣ ع=٠,٥٢٢	م=٠,٢٢٦ ع=٠,١٥٢	م=٠,٠٣٤ ع=٠,٨١٣	م=١,١٩٤ ع=٠,٩٦٤	م=-٠,٥٠٠ ع=٠,٩٥٤	م=٠,٢٧١ ع=٠,١٦٠	م=٠,٠٩٤ ع=٠,٨٦٢

م: متوسط حسابي.

ع: انحراف معياري.

*: تم تثبيتها عند القيمة ٠,٢٠

يُلاحظ من نتائج الجدول بالنسبة لمعلمة القدرة، انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي لها (٠,٠١٠) لوجت في حالة النموذج الثنائي المعدل، كما بلغ (٠,٠١٨) لوجت في حالة النموذج اللوجستي الثلاثي المعلمة، يليه نموذج عتبة التخمين والنموذج القائم على الصعوبة بقيمة بلغت (٠,٠٥٠)لوجت.

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، ففي حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي لمعلمة القدرة (٠,٠٠٥) لوجت، يليه النموذج الثنائي المعدل بقيمة بلغت (٠,٠١٢) لوجت، ثم نموذج عتبة التخمين بقيمة بلغت(٠,٠٢٥) لوجت، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت (٠,٠٦٤) لوجت.

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي لمعلمة القدرة(٠,٠٠٥) لوجت في حالة النموذج اللوجستي الثلاثي المعلمة يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت (٠,٠١١) لوجت، ثم نموذج عتبة التخمين بقيمة بلغت (٠,٠٢٨) لوجت، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت (٠,٠٣٠) لوجت.

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين بلغ المتوسط الحسابي لمعلمة القدرة (٠,٠١١) لوجت في حالة النموذج الثنائي المعدل، يليه النموذج اللوجستي الثلاثي المعلمة بقيمة بلغت (٠,٠١٨) لوجت، ثم نموذج عتبة التخمين بقيمة بلغت (٠,٠٤٨) لوجت، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت قيمته (٠,٠٨٢) لوجت.

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي لمعلمة القدرة (٠,٠١٩) لوجت، وارتفع ظاهريًا ليصل (٠,٠٤٠) لوجت في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي معلمة القدرة (٠,٠٣٢) لوجت، وفي حالة العينة المكونة من (١٥٠٠) مفحوص فقد انخفض ظاهريًا المتوسط الحسابي لمعلمة القدرة ليبلغ القيمة (٠,٠٢٦) لوجت.

كما يُلاحظ من نتائج الجدول بالنسبة لمعلمة الصعوبة انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي لمعلمة الصعوبة (٠,٦٦٩) لوجت في حالة نموذج عتبة التخمين، كما بلغ (٠,٠٦٨) لوجت في حالة النموذج اللوجستي الثلاثي المعلمة،

يليه النموذج القائم على الصعوبة بقيمة بلغت (-٠,٥٢٥) لوجت، ثم النموذج الثنائي المعدل بقيمة بلغت (صفر)لوجت.

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، ففي حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي لمعلمة الصعوبة (-٠,٢٨٢) لوجت، يليه النموذج الثنائي المعدل بقيمة بلغت (-٠,٤٨٢) لوجت، ثم النموذج القائم على الصعوبة بقيمة بلغت (-٠,٥٩٢) لوجت، واخيرًا نموذج عتبة التخمين بقيمة بلغت(-١,٠٧٣) لوجت.

ويُلاحظ من نتائج الجدول بالنسبة لمعلمة الصعوبة انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي لمعلمة الصعوبة(-٠,١٢٢) لوجت في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت (-٠,٣١٥) لوجت، ثم النموذج القائم على الصعوبة بقيمة بلغت(-٠,٥٥٥) لوجت، واخيرًا نموذج عتبة التخمين بقيمة بلغت(-٠,٩٤٤) لوجت.

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين بلغ المتوسط الحسابي لمعلمة الصعوبة في حالة النموذج اللوجستي الثلاثي المعلمة (-٠,٠٩٢) لوجت، يليه النموذج الثنائي المعدل بقيمة بلغت (-٠,١٦٧) لوجت، ثم النموذج القائم على الصعوبة بمتوسط حسابي بلغ(-٠,٥٦٢) لوجت، واخيرًا نموذج عتبة التخمين بقيمة بلغت(-٠,٧٩٧) لوجت.

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي لمعلمة الصعوبة (-٠,٤٨٤) لوجت، وفي حالة الاختبار المكون من (٥٠) فقرة بلغ المتوسط الحسابي لمعلمة الصعوبة(-٠,٨٠٩) لوجت.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي لمعلمة الصعوبة (٠,١٠٦) لوجت، وفي حالة العينة المكونة من (١٥٠٠) مفحوص انخفض ظاهريًا المتوسط الحسابي لمعلمة الصعوبة ليبلغ القيمة (-٠,٦٠٧) لوجت.

ويُلاحظ من نتائج الجدول بالنسبة لمعلمة التمييز انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي لمعلمة التمييز (١,٠٧٠) لوجت في حالة النموذج الثنائي المعدل، كما بلغ (١,٠٣٥) لوجت في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج القائم على الصعوبة بقيمة بلغت (٠,٩٠٨) لوجت، ثم نموذج عتبة التخمين بقيمة بلغت(٠,٧٩٧)لوجت.

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، ففي حالة النموذج اللوجستي القائم على الصعوبة بلغ المتوسط الحسابي لمعلمة التمييز(١,١٤٨) لوجت، يليه النموذج اللوجستي الثلاثي المعلمة بقيمة بلغت (١,٠١٦) لوجت، ثم النموذج الثنائي المعدل بقيمة بلغت(١,٠١٠) لوجت، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت (٠,٨٠٧) لوجت.

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي لمعلمة التمييز(١,٠٢٢) لوجت في حالة النموذج اللوجستي الثلاثي المعلمة يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت (١,٠٤٥) لوجت، ثم النموذج القائم على الصعوبة بقيمة بلغت (١,٠٣١) لوجت، واخيرًا نموذج عتبة التخمين بقيمة بلغت(٠,٧٩٤) لوجت.

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين بلغ المتوسط الحسابي لمعلمة التمييز في حالة النموذج الثنائي المعدل (١,٠٣٣) لوجت، يليه النموذج اللوجستي الثلاثي المعلمة بقيمة بلغت (١,٠٢٩) لوجت، ثم النموذج القائم على الصعوبة بقيمة بلغت (١,٠٢٥) لوجت، واخيرًا نموذج عتبة التخمين بقيمة بلغت (٠,٨١٠) لوجت.

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي لمعلمة التمييز (٠,٩٧٣) لوجت، وفي حالة الاختبار المكون من (٥٠) فقرة بلغ المتوسط الحسابي لمعلمة التمييز(٠,٩٧٤) لوجت.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي معلمة التمييز (٠,٩٥٣) لوجت وفي حال العينة المكونة من (١٥٠٠) مفحوص فقد ارتفع ظاهريًا المتوسط الحسابي لمعلمة التمييز ليبلغ القيمة (٠,٩٩٥) لوجت.

ويُلاحظ من نتائج الجدول بالنسبة لمعلمة التخمين انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي لمعلمة التخمين(٠,٢٢٩) لوجت في حالة النموذج القائم على الصعوبة، كما بلغ المتوسط الحسابي (٠,٢١٨) لوجت في حالة نموذج عتبة التخمين، يليه النموذج اللوجستي الثلاثي المعلمة بمتوسط حسابي بلغ (٠,١٨٢) لوجت.

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، ففي حالة نموذج عتبة التخمين بلغ المتوسط الحسابي لمعلمة التخمين(٠,٢٥١) لوجت، يليه النموذج القائم على الصعوبة بقيمة بلغت (٠,٢٤٩) لوجت، ثم النموذج اللوجستي الثلاثي المعلمة بقيمة بلغت (٠,٢٢٦)لوجت.

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي لمعلمة التخمين (٠,٢٣٦) في حالة النموذج القائم على الصعوبة، يليه نموذج عتبة التخمين بمتوسط حسابي (٠,٢١٩) ، ثم النموذج اللوجستي الثلاثي المعلمة بقيمة بلغت (٠,٢١٦).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين بلغ المتوسط الحسابي لمعلمة التخمين في حالة النموذج القائم على الصعوبة (٠,٢٤٩) ، يليه عتبة التخمين بقيمة بلغت (٠,٢٤٣) ، ثم النموذج اللوجستي الثلاثي المعلمة بقيمة بلغت (٠,١٩٣).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي لمعلمة التخمين(٠,٢٢٤) ، وفي حالة الاختبار المكون من (٥٠) فقرة بلغ المتوسط الحسابي لمعلمة الصعوبة(٠,٢٢٨).

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي لمعلمة التخمين(٠,٢١٠) وفي حال العينة المكونة من (١٥٠٠) مفحوص فقد ارتفع ظاهريًا المتوسط الحسابي لمعلمة التخمين ليبلغ (٠,٢٤٢).

واستكمالاً للإجابة عن هذا السؤال تم استخراج متوسطات الخطأ المعياري للتقدير Standard Error of Estimate (SEE) لمعالم الفقرات (معلمة التمييز ومعلمة الصعوبة ومعلمة التخمين) ومعلمة القدرة للمفحوصين بإختلاف متغيرات الدراسة الأخرى (حجم العينة وطول الاختبار)، إذ يُعتبر هذا المؤشر الإحصائي من مؤشرات قياس دقة التقدير لمعالم الفقرات والأفراد في إطار نظرية الاستجابة للفقرة، فكلما قلّت قيمة هذا المؤشر زادت الدقة في تقدير المعالم، والجدول (١٢) يُبين مقدار قيم متوسطات الخطأ المعياري للتقدير التي تم التوصل إليها لمعالم الفقرات، ولمعلمة القدرة للمفحوصين باختلاف النموذج المُستخدم لمعالجة التخمين وحجم عينة المفحوصين وطول الاختبار.

جدول رقم (١٣)

المتوسطات الحسابية للخطأ المعياري لتقدير معالم الفقرات ومعلمة القدرة المفحوصين عند تطبيق كل نموذج من النماذج المستخدمة لمعالجة التخمين تبعًا لمتغيري الدراسة حجم العينة وطول الاختبار.

حجم العينة	النموذج	طول الاختبار: اختبار (أ) ٢٥				اختبار (ب) ٥٠			
		القدرة	**التمييز**	**الصعوبة**	**التخميين**	**القدرة**	**التمييز**	**الصعوبة**	**التخمين**
٥٠٠ مفحوص	**الثلاثي**	٠,٢٨٥	٠,٠٩٥	٠,٠٦٠	٠,٠٧٦	٠,١٣٢	٠,٠٥٩	٠,٠٢٢	٠,٠٥٤
	الثنائي المعدل	٠,٢٩٣	٠,١١٦	٠,٠٢٢	*	٠,١٣٢	٠,١٢٠	٠,٠٢٠	*
	عتبة التخمين	٠,٣٩٥	٠,١٦٨	٠,١٣٧	٠,٠٨٣	٠,١٣١	٠,١٦٢	٠,١٣٢	٠,٠٧٨
	القائم على الصعوبة	٠,٣٩٦	٠,٣٠٧	٠,١٨٨	٠,١٠٩	٠,٢٨٥	٠,١٥٧	٠,١٣٨	٠,٠٧٧
١٥٠٠ مفحوص	**الثلاثي**	٠,٢٣٧	٠,١٢٠	٠,٠٢٦	٠,٠٥٨	٠,١٥٣	٠,٠٧٣	٠,٠١٦	٠,٠٣٠
	الثنائي المعدل	٠,٢٤٤	٠,١١٨	٠,٠٢٤	*	٠,١٥٩	٠,١١٥	٠,٠٢٠	*
	عتبة التخمين	٠,٢٤١	٠,١٤٣	٠,١٣٠	٠,٠٨٣	٠,١٦٠	٠,١٣٧	٠,٠٩٤	٠,٠٧٩
	القائم على الصعوبة	٠,٣٠٢	٠,١٤٦	٠,١٢٥	٠,٠٧٥	٠,١٨٦	٠,١٣٣	٠,١١٥	٠,٠٦٨

* تم تثبيت معلمة التخمين عند القيمة ٠,٢٠

يُلاحظ من نتائج الجدول (١٣) انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة القدرة (٠,٢٠٨) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعدل بقيمة بلغت (٠,٢١٢)، يليه نموذج عتبة التخمين بقيمة بلغت (٠,٢٦٣)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٣٤٠).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ الانخفاض الظاهري لجميع قيم المتوسطات الحسابية للخطأ المعياري لتقدير معلمة القدرة، ففي حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,١٩٥)، يليه النموذج الثنائي المعدل ونموذج عتبة التخمين بقيمة بلغت(٠,٢٠١)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت (٠,٢٤٤).

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة القدرة(٠,٢٦١) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت (٠,٢٦٩)، يليه نموذج عتبة التخمين بقيمة بلغت(٠,٣١٨)، ثم النموذج القائم على الصعوبة بقيمة بلغت(٠,٣٤٩).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة القدرة (٠,١٤٢) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه نموذج عتبة التخمين والنموذج الثنائي المعدل بقيمة بلغت (٠,١٤٦)، واخيرًا النموذج القائم على الصعوبة بمتوسط حسابي بلغ(٠,٢٣٦).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة القدرة (٠,٢٩٩)، وانخفض ظاهريًا ليصل (٠,١٦٧) في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة القدرة (٠,٢٥٦) وفي حال العينة المكونة من (١٥٠٠) مفحوص فقد انخفض ظاهريًا المتوسط الحسابي للخطأ المعياري لتقدير معلمة القدرة ليبلغ القيمة (٠,٢١٠).

يُلاحظ من نتائج الجدول (١٣) بالنسبة لمعلمة التمييز انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التمييز (٠,٠٧٧) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعدل بقيمة بلغت (٠,١١٨)، يليه نموذج عتبة التخمين بقيمة بلغت(٠,١٦٥)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٢٣٢).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ أن قيم المتوسطات الحسابية للخطأ المعياري للتقدير لمعلمة التمييز قد جاءت كما يلي: في حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,٠٩٧)، يليه النموذج الثنائي المعدل بقيمة بلغت(٠,١١٧) ثم النموذج القائم على الصعوبة بقيمة بلغت(٠,١٣٩)، واخيرًا نموذج عتبة التخمين بقيمة بلغت(٠,١٤٠).

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التمييز(٠,١٠٨) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعدل بقيمة بلغت(٠,١١٧)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,١٥٦)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٢٢٧). وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يُلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التمييز (٠,٠٦٦) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعدل بقيمة بلغت(٠,١١٨)، ثم النموذج القائم على الصعوبة بقيمة بلغت(٠,١٤٥)، واخيرًا نموذج عتبة التخمين بقيمة بلغت(٠,١٥٠)

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التمييز (٠,١٥٢)، وانخفض ظاهريًا ليصل (٠,١٢٠) في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التمييز (٠,١٤٨) وفي حالة العينة المكونة من (١٥٠٠) بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التمييز (٠,١٢٣).

كما يُلاحظ من نتائج الجدول بالنسبة لمعلمة الصعوبة انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة الصعوبة (٠,٠٢١) في حالة النموذج الثنائي المعدل، كما بلغ (٠,٠٤١) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه نموذج عتبة التخمين بقيمة بلغت(٠,١٣٥)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,١٦٣).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يلاحظ أن قيم المتوسطات الحسابية للخطأ المعياري للتقدير لمعلمة الصعوبة قد جاءت كما يلي: في حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,٠٢١)، يليه النموذج الثنائي المعدل بقيمة بلغت(٠,٠٢٢) ثم نموذج عتبة التخمين بقيمة بلغت (٠,١١٢)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت (٠,١٢٠).

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة الصعوبة (٠,٠٤٣)، في حالة النموذج اللوجستي الثلاثي المعلمة يليه النموذج الثنائي المعدل بقيمة بلغت(٠,٠٢٣)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,١٣٤)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,١٥٧).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة الصعوبة (٠,٠١٩) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعدل بمتوسط بلغ(٠,٠٢٠)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,١١٣)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,١٢٧).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة الصعوبة (٠,٠٨٩)، وانخفض ظاهريًا ليصل (٠,٠٦٩) في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة الصعوبة (٠,٠٩٠) وفي حال العينة المكونة من (١٥٠٠) مفحوص فقد انخفض ظاهريًا المتوسط الحسابي للخطأ المعياري لتقدير معلمة الصعوبة ليبلغ القيمة (٠,٠٦٨)، وبالنسبة لمعلمة التخمين يُلاحظ من نتائج الجدول انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التخمين (٠,١١٠) في حالة النموذج اللوجستي الثلاثي المعلمة، كما بلغ (٠,١٨٠) في حالة نموذج عتبة التخمين، يليه النموذج القائم على الصعوبة بمتوسط حسابي بلغ(٠,٢٤٠).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يلاحظ أن قيم المتوسطات الحسابية للخطأ المعياري للتقدير لمعلمة التخمين قد جاءت كما يلي: في حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,١١٠)، يليه نموذج عتبة التخمين بقيمة بلغت(٠,١٤٠)، ثم النموذج القائم على الصعوبة بقيمة بلغت(٠,١٧٠). كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التخمين (٠,١١٠)، في حالة النموذج اللوجستي الثلاثي المعلمة، يليه نموذج عتبة التخمين بمتوسط حسابي بلغ (٠,١٨٠)، واخيرًا نموذج عتبة التخمين بمتوسط حسابي بلغ(٠,٢٣٠).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يُلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التخمين (٠,١٠٠) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه نموذج عتبة التخمين بمتوسط حسابي بلغ(٠,١٤٠)، واخيرًا النموذج القائم على الصعوبة بمتوسط حسابي بلغ(٠,١٨٠).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التخمين (٠١٧٦)، وانخفض ظاهريًا ليصل (٠,١٤٠) في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي للخطأ المعياري لتقدير معلمة التخمين (٠,١٧٣) وفي حالة العينة المكونة من (١٥٠٠) مفحوص فقد انخفض ظاهريًا المتوسط الحسابي للخطأ المعياري لتقدير معلمة التخمين ليبلغ القيمة (٠,١٣٦).

أولاً: النتائج المتعلقة بالإجابة عن سؤال الدراسة الأول:

للإجابة عن السؤال الأول للدراسة والذي يدور حول فيما إذا كان هنالك اختلاف في دقة تقدير معلمة القدرة باختلاف النموذج المستخدم لمعالجة التخمين في إطار النموذج اللوجستي الثلاثي المعلمة وتبعًا لمتغيري طول الاختبار (٢٥ فقرة و٥٠ فقرة) وعدد المفحوصين (٥٠٠ مفحوص و١٥٠٠ مفحوص) تم استخراج المتوسطات الحسابية والانحرافات المعيارية للجذر التربيعي لمربع الخطأ في تقدير معلمة القدرة، ويبين الجدول (١٤) هذه المتوسطات والانحرافات المعيارية تبعًا لمتغيري طول الاختبار وعدد الحالات المولدة.

جدول (١٤)

المتوسطات الحسابية لمؤشر الجذر التربيعي لمتوسط مربع الخطأ (RMSE) لتقدير معلمة القدرة والانحرافات المعيارية المناظرة لها في حالة النماذج الأربعة المختلفة المسُتخدمة لمعالجة التخمين تبعًا لحجم عينة البيانات المولدة وعدد فقرات الاختبار.

حجم العينة	النموذج	طول الاختبار			
		٢٥ فقرة		٥٠ فقرة	
		المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
٥٠٠ مفحوص	الثلاثي	٠,٢٨٨	٠,٥٢٠	٠,١٣٢	٠,١٩٠
	الثنائي المعدل	٠,٢٩٥	٠,٥٢٩	٠,١٣٣	٠,١٩٠
	عتبة التخمين	٠,٢٩٩	٠,٥٦٢	٠,١٣٧	٠,٢٠٤
	القائم على الصعوبة	٠,٤٠٠	٠,٦٨٦	٠,٢٢٨	٠,٣٤٥
١٥٠٠ مفحوص	الثلاثي	٠,٢٣٨	٠,٣٥٢	٠,١٥٤	٠,٢٤٤
	الثنائي المعدل	٠,٢٤٥	٠,٣٦٦	٠,١٦٣	٠,٢٥٩
	عتبة التخمين	٠,٢٤١	٠,٣٦٢	٠,١٦٢	٠,٢٥٨
	القائم على الصعوبة	٠,٣٠٦	٠,٤٥٧	٠,٢٠٩	٠,٣٣٧

يُلاحظ من نتائج الجدول انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة (٠,٢١٠) في حالة النموذج اللوجستي الثلاثي المعلمة، كما بلغ (٠,٢١٤) في حالة النموذج الثنائي المعدل، يليه نموذج عتبة التخمين بقيمة بلغت(٠,٢١٨)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٣١٤).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ الانخفاض الظاهري لجميع قيم الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة، ففي حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,١٩٦)، يليه نموذج عتبة التخمين

بقيمة بلغت(٠,٢٠١)، ثم النموذج الثنائي المعدل بقيمة بلغت(٠,٢٠٤)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٢٥٧).

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة(٠,٢٦٣) في حالة النموذج اللوجستي الثلاثي المعلمة يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت(٠,٢٧٠)، ثم نموذج عتبة التخمين بقيمة بلغت (٠,٢٧١)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت (٠,٣٥٣).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين، يُلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة (٠,١٤٣) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت(٠,١٤٨) ثم نموذج عتبة التخمين بقيمة بلغت(٠,١٤٩)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٢١٩). وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة (٠,٢٨٩)، وانخفض ظاهريًا ليصل (٠,١٦٥) في حالة الاختبار المكون من (٥٠) فقرة. وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة (٠,٢٣٩)، وفي حال العينة المكونة من (١٥٠٠) مفحوص فقد انخفض ظاهريًا المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ ليبلغ القيمة (٠,٢١٥).

وللوقوف على أن التغييرات في قيم متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة تختلف تبعًا لمتغيرات الدراسة: النموذج المستخدم لمعالجة التخمين وطول الاختبار وحجم عينة المفحوصين المولدة لهم معلمة القدرة والتفاعلات بينها، جرى استخدام تحليل التباين الثلاثي للقياسات المتكررة لتأثير هذه المتغيرات والتفاعل بينها على المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة والجدول رقم(١٥) يبين ملخص نتائج تحليل التباين.

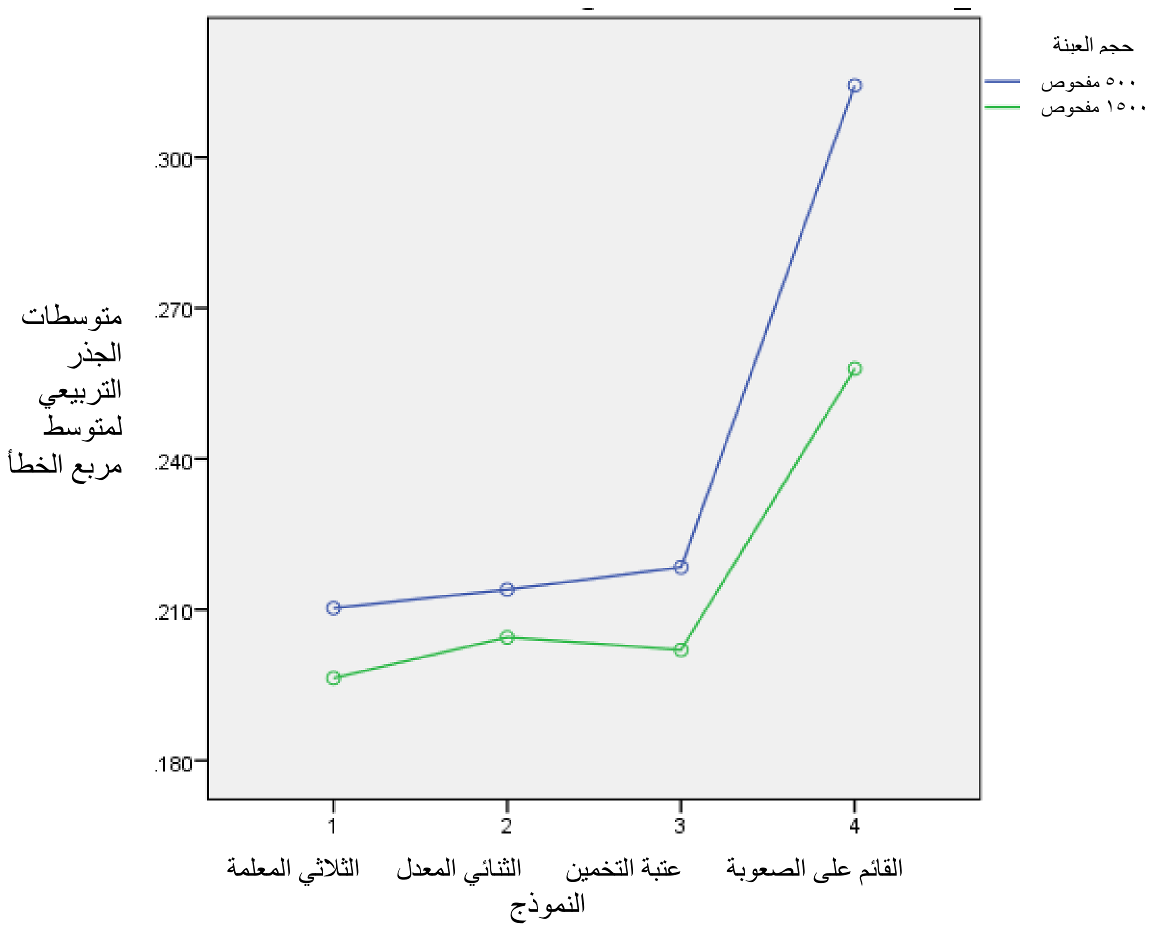
الجدول (١٥)

ملخص نتائج تحليل التباين الثلاثي للقياسات المتكررة لأثر متغيرات الدراسة: النموذج المستخدم لمعالجة التخمين وطول الاختبار وحجم العينة والتفاعلات بينها على متوسطات الجذر التربيعي لمتوسطات مربع الخطأ في تقدير معلمة القدرة للمفحوصين.

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	الدلالة الإحصائية
داخل الأفراد	النموذج.	١٣,٩٦١	٣	٤,٦٥٤	٢٤٦,٢٧٩	*٠,٠٠٠
	النموذج* عدد الفقرات.	٠,١١١	٣	٠,٠٣٧	١,٩٦٦	٠,١١٧
	النموذج*حجم العينة.	١,٠٦١	٣	٠,٣٥٤	١٨,٧١٨	*٠,٠٠٠
	النموذج*عدد الفقرات*حجم العينة.	٠,٠١٥	٣	٠,٠٠٥	٠,٢٦٣	٠,٨٥٢
	الخطأ	٢٢٦,٥١٧	١١٩٨٨	٠,٠١٩		
بين الأفراد	عدد الفقرات	٤٦,٠٦٥	١	٤٦,٠٦٥	٩٦,١٢١	*٠,٠٠٠
	حجم العينة	١,٧٣٦	١	١,٧٣٦	٣,٦٢٣	٠,٠٥٧
	عدد الفقرات × حجم العينة	٤,٤٩٣	١	٤,٤٩٣	٩,٣٧٦	*٠,٠٠٢
	الخطأ	١٩١٥,٠٣٢	٣٩٩٦	٠,٤٧٩		

يُلاحظ من نتائج الجدول (١٥) عدم وجود أثرٍ ذي دلالة إحصـــائية للتفاعل الثلاثي بين كلّ من النموذج المستخدم لمعالجة التخمين، وعدد الفقرات وحجم العينة على متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة، إذ كانت قيمة (ف) غير دالة إحصائيًا عند مستوى الدلالة($\alpha = 0.05$)؛ ممّا يُشير إلى عدم وجود أثرٍ ذي دلالة إحصائية للتفاعلات الثنائية للمتغيرات تبعًا لمستويات المتغير الثالث، كذلك يُلاحظ عدم وجود أثرٍ ذي دلالة إحصائية للتفاعل الثنائي بين النموذج المستخدم لمعالجة التخمين وطول الاختبار على تقدير متوسطات الجذر التربيعي لمتوسط مربع الخطأ لمعلمة القدرة للمفحوصين، فقد كانت قيم الدلالة الخاصة بـــ (ف) تزيد عن (٠,٠٥) ممّا يُشير إلى عدم اختلاف فاعلية النماذج المستخدمة باختلاف عدد فقرات الاختبار، بينما يُلاحظ وجود أثرٍ ذي دلالة إحصائية للتفاعل الثنائي بين النموذج المستخدم لمعالجة التخمين وحجم العينة،

على تقدير متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة والشكل رقم(٥) يوضح هذا التفاعل.

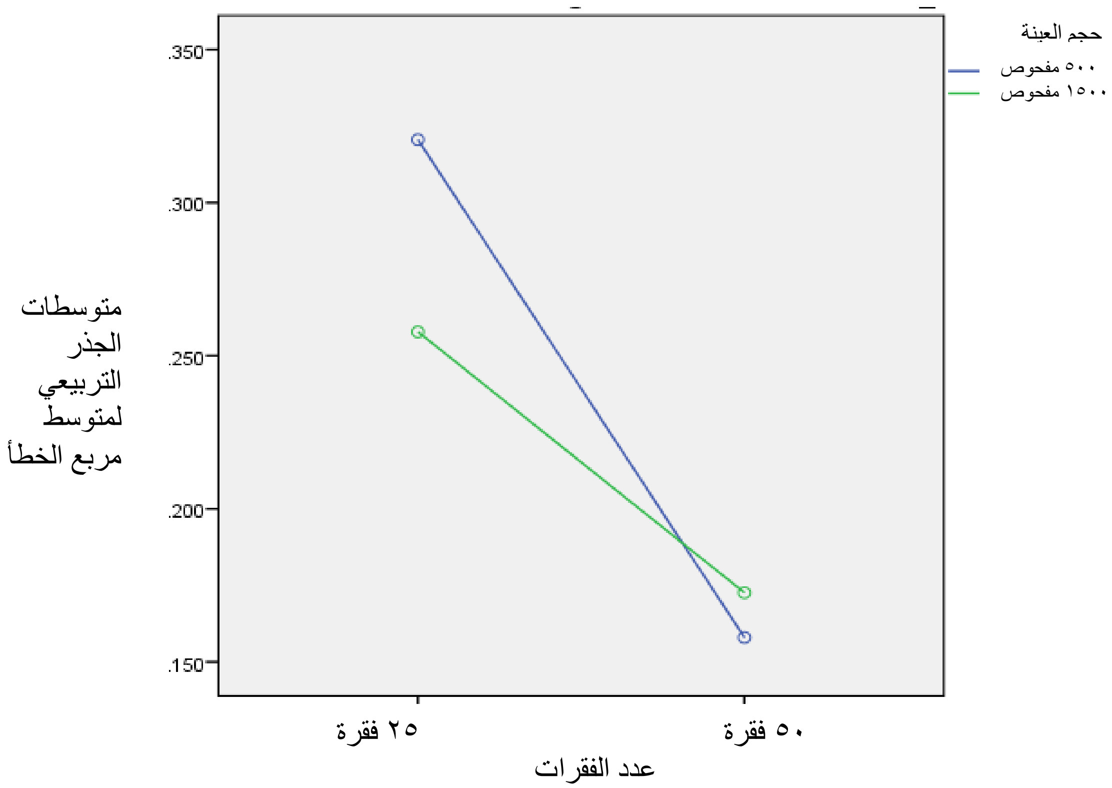


شكل (٥)
التمثيل البياني للتفاعل الثنائي بين النموذج المستخدم لمعالجة التخمين وحجم عينة المفحوصين(٥٠٠ مفحوص و ١٥٠٠ مفحوص) على متوسطات مؤشر الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة.

يُلاحظ من الشكل (٥) الاختلاف الواضح في قيم المتوسطات الحسابية لمؤشر الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة تبعًا لحجم البيانات المولدة، حيث يُلاحظ عمومًا أن قيم المتوسطات أعلى منها في حالة البيانات للعينة المكونة من (٥٠٠) مفحوص منها بحالة البيانات للعينة المكونة من (١٥٠٠) مفحوص، كذلك يُلاحظ التباين في قيم متوسطات الجذر التربيعي لمتوسط مربع الخطأ لمعلمة القدرة تبعًا للطريقة، حيث يُلاحظ أن الفروق لقيم الجذر التربيعي لمتوسط مربع الخطأ لمعلمة القدرة تقريبًا منتظمة ومتقاربة في حالة النموذج اللوجستي الثلاثي المعلمة والنموذج اللوجستي الثنائي المعدل، ونموذج عتبة التخمين، لكن يُلاحظ الفرق في قيم متوسطات الجذر التربيعي لمتوسط مربع الخطأ لمعلمة القدرة في حالة كل من العينة المكونة من بيانات مولدة لـ(٥٠٠)

مفحوص والعينة المكونة من البيانات المولدة لـ(١٥٠٠) مفحوص؛ الأمر الذي يُشير إلى تأثير حجم العينة على تقدير هذا المؤشر في حالة النموذج القائم على الصعوبة بدرجة اكبر مقارنة بالنماذج الأخرى المستخدمة لمعالجة التخمين في إطار هذه الأطروحة.

كذلك يُلاحظ من نتائج الجدول وجود أثر ذي دلالة إحصائية للتفاعل بين عدد الفقرات (٢٥ فقرة و ٥٠ فقرة) وحجم عينة المفحوصين (٥٠٠ مفحوص و ١٥٠٠ مفحوص) على متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة، حيث كانت قيمة (ف) دالة إحصائيًا عند مستوى الدلالة (α=0.05) والشكل (٦) يوضح هذا التفاعل.



شكل رقم (٦)

التمثيل البياني للتفاعل الثنائي بين حجم عينة المفحوصين (٥٠٠ مفحوص و١٥٠٠ مفحوص) وعدد الفقرات (٢٥ فقرة و٥٠ فقرة) على متوسطات مؤشر الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة.

يُلاحظ من الشكل (٦) أن قيم متوسطات مؤشر الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة القدرة للمفحوصين تختلف باختلاف كلّ من: حجم العينة وطول الاختبار، حيث يُلاحظ أن الزيادة في عدد الفقرات يُصاحبه انخفاضُ واضحُ في حجم قيمة هذا المؤشر بغض النظر عن حجم عينة البيانات المولدة، في حين يُلاحظ ان حجم التأثير يكون أكبر في حالة البيانات المولدة لعينة مكونة من (٥٠٠) مفحوص وطول اختبار (٥٠) فقرة.

واستكمالاً للإجابة عن هذا السؤال تم استخراج المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة القدرة كما في الجدول الآتي.

جدول (١٦)

المتوسطات الحسابية لمؤشر التحيز (Bias) في تقدير معلمة القدرة (المُقدرة) للمفحوصين والانحرافات المعيارية المناظرة لها تبعًا لحجم عينة البيانات المولدة وعدد فقرات الاختبار.

حجم العينة	النموذج	طول الاختبار			
		٢٥ فقرة		٥٠ فقرة	
		المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
٥٠٠ مفحوص	الثلاثي	٠,٠٤٠	٠,٠١٩	٠,٠٠٣	٠,٠٢٠
	الثنائي المعدل	٠,٠٣٣	٠,٠٢١	٠,٠١٤	٠,٠٢٣
	عتبة التخمين	٠,٠٦٤	٠,٠٢٠	٠,٠٤٠	٠,٠٢١
	القائم على الصعوبة	٠,٠٥٣	٠,٠٣٠	٠,٠٤٣	٠,٠٢٠
١٥٠٠ مفحوص	الثلاثي	٠,٠١٢	٠,٠١٢	٠,٠٠٢	٠,٠١٢
	الثنائي المعدل	٠,٠٢٦	٠,٠١٦	٠,٠٠١	٠,٠١٦
	عتبة التخمين	٠,٠٣٤	٠٠,٠١٣	٠,٠٢١	٠,٠٢١
	القائم على الصعوبة	٠,٠٤٨	٠,٠٧٠	٠,٠٩٤	٠,٠٣٢

ويُلاحظ من نتائج الجدول انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة القدرة (٠,٠٢٢) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعدل بقيمة بلغت(٠,٠٢٤)، ثم النموذج القائم على الصعوبة بقيمة بلغت (٠,٠٤٨)، واخيرًا نموذج عتبة التخمين بقيمة بلغت(٠,٠٥٢).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ الانخفاض الظاهري لجميع قيم المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة القدرة، ففي

حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,٠٠٧)، يليه النموذج الثنائي المعدل بقيمة بلغت(٠,٠١٤)، ثم النموذج نموذج عتبة التخمين بقيمة بلغت(٠,٠٢٨)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٧١).

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة القدرة (٠,٠٢٦) في حالة النموذج اللوجستي الثلاثي المعلمة يليه النموذج الثنائي المعلمة المُعدل بقيمة بلغت(٠,٠٣٠)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,٠٤٧)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٥١).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يلاحظ أن جميع قيم متوسطات التحيز في تقدير معلمة القدرة قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة القدرة (٠,٠٠٣) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت(٠,٠٠٨)، ثم نموذج عتبة التخمين بقيمة بلغت (٠,٠٣١)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٦٩).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة القدرة (٠,٠٣٩)، وانخفض ظاهريًا ليصل (٠,٠٢٨) في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة القدرة (٠,٠٣٧) وفي حالة العينة المكونة من (١٥٠٠) مفحوص فقد انخفض المتوسط الحسابي لمؤشر التحيز في تقدير معلمة القدرة ظاهريًا ليبلغ القيمة (٠,٠٣٠).

وللوقوف على أن التغييرات في قيم المتوسـطات الحسـابية لمؤشـر التحيز في تقدير معلمة القدرة تختلف تبعًا لمتغيرات الدراسـة: النموذج المسـتخدم لمعالجة التخمين وطول الاختبار وحجم عينة المفحوصـين المولدة لهم معلمة القدرة والتفاعلات بينها، جرى اسـتخدام تحليل التباين الثلاثي للقياسات المتكررة لتأثير هذه المتغيرات والتفاعل بينها على المتوسطات الحسابية لمؤشر التحيز لتقدير معلمة القدرة والجدول (١٧) يُبيّن ملخص نتائج تحليل هذا التباين.

جدول (١٧)

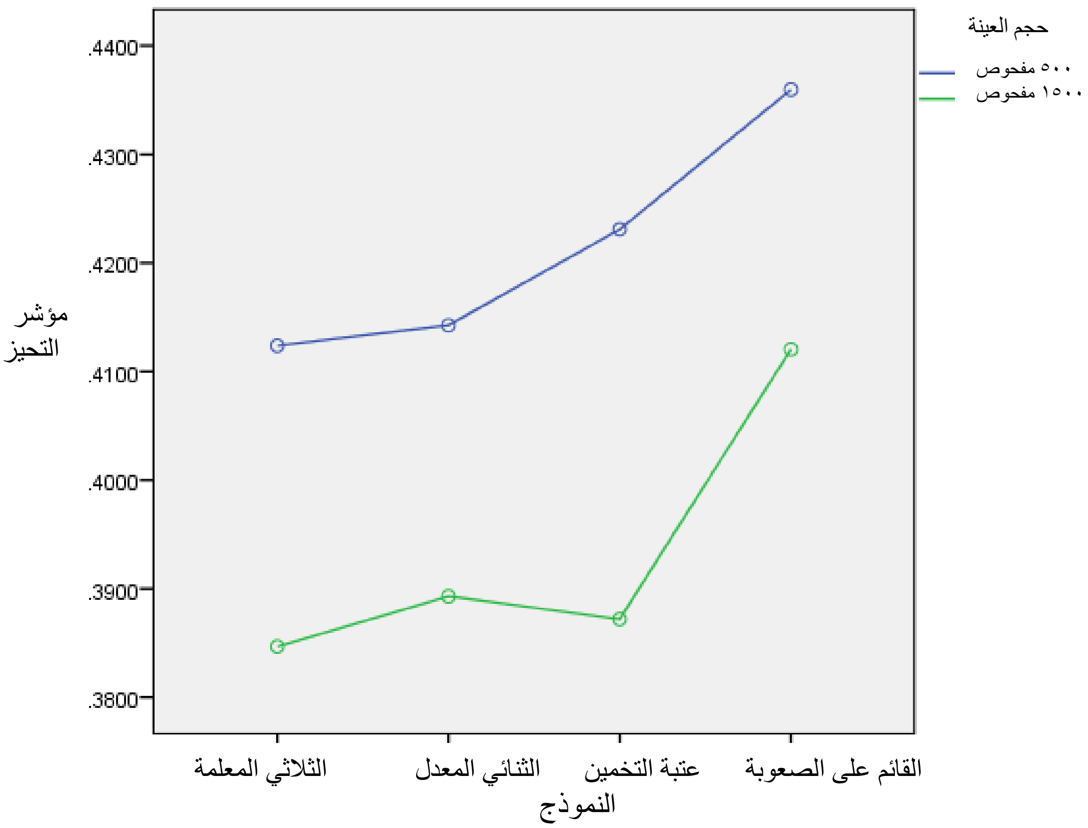
ملخص نتائج تحليل التباين الثلاثي للقياسات المتكررة لأثر متغيرات الدراسة: النموذج المستخدم لمعالجة التخمين وطول الاختبار وحجم العينة والتفاعلات بينها على متوسطات مؤشر التحيز في تقدير معلمة القدرة للمفحوصين.

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة الإحصائية
داخل الأفراد	النموذج.	٥,٠٦٢	٣	١,٦٨٧	٤٢,٦٤٤	٠,٠١١*
	النموذج* عدد الفقرات.	١,٤٥٢	٣	٠,٤٨٤	١٢,٢٣٣	٠,٠٠٣*
	النموذج*حجم العينة.	٠,١٥٣	٣	٠,٠٥١	١,٢٨٨	٠,٠٠٠*
	النموذج*عدد الفقرات*حجم العينة.	٠,٣٣٩	٣	٠,١١٣	٢,٨٥٧	٠,٠٠١*
	الخطأ	٤٧٤,٣٢٤	١١٩٨٨	٠,٠٤٠		
بين الأفراد	عدد الفقرات	٨,٨٦٣	١	٨,٨٦٣	٣١,٩٥٩	٠,٠٠٠*
	حجم العينة	٢,٥٣٦	١	٢,٥٣٦	٩,١٤٤	٠,٠٠٣*
	عدد الفقرات × حجم العينة	٠,٠٠٣	١	٠,٠٠٣	٠,٠١٠	٠,٩٢١
	الخطأ	١١٠٨,١٧٣	٣٩٩٦	٠,٢٧٧		

يُلاحظ من نتائج الجدول (١٧) وجود أثرٍ ذي دلالة إحصائية للتفاعل الثلاثي بين النموذج المُستخدم لمعالجة التخمين وطول الاختبار(عدد الفقرات) وحجم العينة، وللتفاعل الثنائي بين النموذج وحجم العينة وللتفاعل الثنائي بين النموذج وعدد الفقرات على متوسـطات مؤشـر التحيز في تقدير معلمة القدرة للمفحوصين، فقد كانت قيم الدلالة الخاصة بـ(ف) تقل عن (٠,٠٥) عند كل حالة.

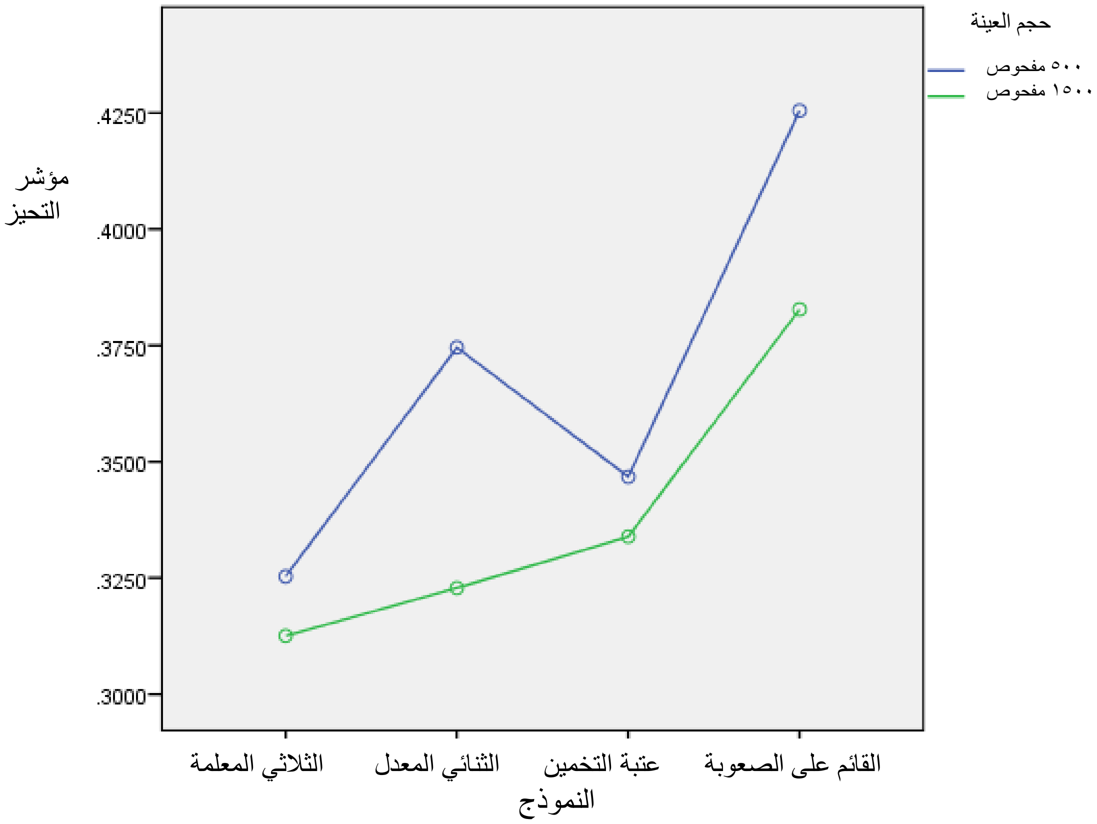
أما بالنسبة للتباين بين الأفراد فُتشير نتائج الجدول إلى وجود أثرٍ دال إحصائيًا لكل من حجم العينة وعدد فقرات الاختبار، فقد كانت الدلالة الخاصـــة بقيمة (ف) أقل من (٠,٠٥) عند كل حالة، بينما يلاحظ من الجدول عدم وجود أثرٍ للتفاعل الثنائي بين عدد الفقرات وحجم عينة المفحوصـــين فقد كانت قيم الدلالة الخاصة بـ(ف) تزيد عن (٠,٠٥) عند كل حالة.

ولدراسة التفاعل الثلاثي بين كل من النموذج المستخدم لمعالجة التخمين وحجم عينة المفحوصين ومستويات متغير عدد الفقرات للاختبار على متوسطات مؤشر التحيز في تقدير معلمة القدرة تم استخراج التمثيل البياني كما في الشكلين (٧) و(٨).



شكل رقم (٧)

التمثيل البياني للتفاعل الثلاثي بين كل من النموذج المستخدم لمعالجة التخمين وحجم العينة عندما يكون طول الاختبار (٢٥ فقرة) على متوسطات مؤشر التحيز في تقدير معلمة القدرة.



شكل رقم (٨)
التمثيل البياني للتفاعل الثلاثي بين كل من النموذج المستخدم لمعالجة التخمين وحجم العينة عندما يكون طول الاختبار (٥٠ فقرة) على متوسطات مؤشر التحيز في تقدير معلمة القدرة.

يُلاحظ من الشكلين (٧) و (٨) أن التفاعل الثنائي بين تأثير النموذج المستخدم لمعالجة التخمين وحجم عينة البيانات المولدة (٥٠٠ مفحوص و١٥٠٠ مفحوص) على متوسطات مؤشر التحيز في تقدير معلمة القدرة للمفحوصين يختلف باختلاف عدد الفقرات المكونة للاختبار، حيث يُلاحظ أن فاعلية النماذج المستخدمة لمعالجة التخمين كما يدل عليها مؤشر التحيز عمومًا أكبر في حالة أن يكون حجم العينة (١٥٠٠) مفحوص منها في حالة العينة المكونة من (٥٠٠) مفحوص، لكن يُلاحظ الانتظام في التزايد في قيم مؤشر التحيز في حالة الاختبار المكون من (٢٥) فقرة عند الانتقال من النموذج اللوجستي الثلاثي المعلمة، ثم النموذج الثنائي المعدل ثم نموذج عتبة التخمين ثم النموذج القائم على الصعوبة لكن يبدو هذا متذبذبًا في حالة أن يكون الاختبار مكونًا من (٥٠) فقرة ويبدو هذا الأمر واضحًا في حالة نموذج عتبة التخمين.

النتائج المتعلقة بالإجابة عن سؤال الدراسة الثاني:

هل تختلف دقة تقدير معالم الفقرات (معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين) باختلاف النموذج المستخدم لمعالجة التخمين في إطار النموذج اللوجستي الثلاثي المعلمة تبعًا لمتغيري طول الاختبار (٢٥ فقرة و٥٠ فقرة) وعدد المفحوصين (٥٠٠ مفحوص و١٥٠٠ مفحوص).

للإجابة عن هذا السؤال تم إيجاد الجذر التربيعي لمتوسط مربع الخطأ لكل معلم من معالم الفقرة (معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين) على التوالي، حيث تزداد دقة التقديرات كلما اقتربت قيمة هذا المؤشر من الصفر.

أ- **معلمة التمييز:**

يُبين الجدول (١٨) المتوسطات الحسابية والانحرافات المعيارية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز باختلاف النموذج المستخدم لمعالجة التخمين وحجم عينة المفحوصين (٥٠٠ مفحوص و ١٥٠٠ مفحوص) وطول الاختبار (٢٥ فقرة و٥٠ فقرة).

جدول (١٨)

المتوسطات الحسابية لمؤشر الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز والانحرافات المعيارية المناظرة لها في حالة النماذج الأربعة المختلفة المستخدمة لمعالجة التخمين تبعًا لحجم عينة البيانات المولدة وعدد فقرات الاختبار.

حجم العينة	النموذج	طول الاختبار			
		٢٥ فقرة		٥٠ فقرة	
		المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
٥٠٠ مفحوص	الثلاثي	٠,١٠٠	٠,١٨٤	٠,٠٦٠	٠,٠٩٥
	الثنائي المعدل	٠,١٢٦	٠,٢٦٩	٠,١٢٢	٠,٢٥٤
	عتبة التخمين	٠,١٧٣	٠,٢٧٩	٠,١٦٧	٠,٣٠٨
	القائم على الصعوبة	٠,٣١٢	٠,٦٣٤	٠,١٦٥	٠,٢٦٩
١٥٠٠ مفحوص	الثلاثي	٠,١٢٤	٠,٢١٨	٠,٠٧٣	٠,١٢٤
	الثنائي المعدل	٠,١٢٣	٠,٢٠٢	٠,١١٦	٠,١٧٢
	عتبة التخمين	٠,١٤٨	٠,٢٢٨	٠,١٤٢	٠,٢٤١
	القائم على الصعوبة	٠,١٥٢	٠,٢٢٧	٠,١٣٠	٠,١٩٩

يُلاحظ من نتائج الجدول انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز (٠,٠٧٩) في حالة النموذج اللوجستي الثلاثي المعلمة، كما بلغ (٠,١٢٤) في حالة النموذج الثنائي المعدل، يليه نموذج عتبة التخمين بقيمة بلغت(٠,١٧٠) واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٢٣٩).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ الانخفاض الظاهري لجميع قيم متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز، ففي حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,٠٩٩)، يليه النموذج الثنائي المعدل بمتوسط حسابي بلغ(٠,١١٩)، ثم النموذج القائم على الصعوبة بمتوسط حسابي بلغ(٠,١٤١)، واخيرًا نموذج عتبة التخمين بمتوسط حسابي بلغ (٠,١٤٥)

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز (٠,١١١) في حالة النموذج اللوجستي الثلاثي المعلمة يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت(٠,١٢٥)، ثم نموذج عتبة التخمين بقيمة بلغت (٠,١٦١)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت (٠,٢٣٢).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز (٠,٠٦٧) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعدل بقيمة بلغت(٠,١١٤) ثم النموذج القائم على الصعوبة بقيمة بلغت (٠,١٤٨)، واخيرًا نموذج عتبة التخمين بقيمة بلغت (٠,١٥٥).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز (٠,١٥٧)، وانخفض ظاهريًا ليصل (٠,١٢٢) في حالة الاختبار المكون من (٥٠) فقرة. وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز (٠,١٥٣)، وفي حالة العينة المكونة (١٥٠٠) مفحوص فقد انخفض ظاهريًا المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز ليبلغ القيمة (٠,١٢٦).

وللوقوف على أن التغييرات في قيم متوسطات الجذر التربيعي لمربعات الخطأ في تقدير معلمة التمييز تختلف تبعًا لمتغيرات الدراسة: نموذج معالجة التخمين وطول الاختبار وحجم عينة المفحوصين المولدة لهم معلمة التمييز والتفاعلات بينها، جرى استخدام نتائج تحليل التباين الثلاثي للقياسات المتكررة لتأثير هذه المتغيرات والتفاعل بينها على المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز، والجدول (١٩) يُبين ملخص نتائج هذا التحليل.

جدول (١٩)

ملخص نتائج تحليل التباين الثلاثي للقياسات المتكررة لأثر متغيرات الدراسة: النموذج المستخدم لمعالجة التخمين وطول الاختبار وحجم العينة والتفاعلات بينها على متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز للفقرات.

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة الإحصائية
داخل الفقرات	**النموذج**	٠,٧٦٢	٣	٠,٢٥٤	٦,٦٤٥	*٠,٠٠٠
	النموذج* عدد الفقرات	٠,١٤١	٣	٠,٠٤٧	١,٢٢٦	٠,٣٠٠
	النموذج*حجم العينة	٠,٢٥٥	٣	٠,٠٨٥	٢,٢٢٤	٠,٠٨٥
	النموذج*عدد الفقرات*حجم العينة	٠,١٠٧	٣	٠,٠٣٦	٠,٩٣٠	٠,٤٢٦
	الخطأ	١٦,٧٤٨	٤٣٨	٠,٠٣٨		
بين الفقرات	**عدد الفقرات**	٠,١٠٠	١	٠,١٠٠	٠,١٨٦	٠,٦٦٧
	حجم العينة	٠,٥٧٥	١	٠,٥٧٥	١,٠٧٢	٠,٣٠٢
	عدد الفقرات × حجم العينة	٠,٠٧٩	١	٠,٠٧٩	٠,١٤٨	٠,٧٠١
	الخطأ	٧٨,٣٣٥	١٤٦	٠,٥٣٧		

يُلاحظ من نتائج الجدول (١٩) عدم وجود أثرٍ للتفاعلات الثلاثية والثنائية لمتغيرات الدراسة لأثر النموذج المستخدم لمعالجة التخمين، وعدد الفقرات المولد لهم بيانات على تقدير الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز حيث كانت جميع قيم (ف) ليس لها دلالة إحصائية عند مستوى الدلالة (α=0.05)، بينما يُلاحظ وجود أثرٍ دال إحصائيًا يُعزى للنموذج المستخدم لمعالجة التخمين حيث كانت قيم الدلالة الخاصة بـ(ف) أقل من (٠,٠٥)، أما بالنسبة للتباين بين الافراد تُشير نتائج الجدول إلى عدم وجود أثرٍ للتفاعل الثنائي و لعدد الفقرات و لعدد المفحوصين المُولد لهم بيانات؛ فقد كانت جميع قيم (ف) لها ليست دالة إحصائيًا عند مستوى الدلالة (α=0.05).

ولمعرفة أي الفروق بين متوسطات قيم الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التمييز تبعًا للنماذج المستخدمة لمعالجة التخمين كان له تأثير دال إحصائيًا فقد تم استخراج نتائج اختبار أقل فرق دال Least Square Differences (LSD) والتي تبدو في الجدول (٢٠).

جدول (٢٠)

نتائج اختبار أقل فرق دال (LSD) للمقارنات الثنائية بين متوسطات الجذر التربيعي لمتوسط مربع الخطأ لتقدير معلمة التمييز باختلاف النموذج المستخدم لمعالجة التخمين.

النموذج (١)	**النموذج (٢)**	**متوسط الفروق بين النموذجين (١- ٢)**	**مستوى الدلالة**
الثلاثي المعلمة	الثنائي المعدل	-٠,٠٩١*	٠,٠٠٧*
	عتبة التخمين	-٠,٠٩٩*	٠,٠٠٠*
	القائم على الصعوبة	-٠,٢٢٢	٠,٠٠٠*
الثنائي المعدل	الثلاثي المعلمة	٠,٠٩١	٠,٠٠٧*
	عتبة التخمين	-٠,٠٠٨	٠,٨٠٩
	القائم على الصعوبة	-٠,١٣١	٠,٠١٠*
عتبة التخمين	الثلاثي المعلمة	٠,٠٩٩	٠,٠٠٠*
	الثنائي المعدل	٠,٠٠٨	٠,٨٠٩
	القائم على الصعوبة	-٠,١٢٣	٠,٠٠٥*
القائم على الصعوبة	الثلاثي المعلمة	٠,٢٢٢	٠,٠٠٠*
	الثنائي المعدل	٠,١٣١	٠,٠١٠*
	عتبة التخمين	٠,١٢٣	٠,٠٠٥*

يُلاحظ من نتائج الجدول (٢٠) أن الفروق بين المتوسـطات الحسـابية للجذر التربيعي لمتوسـط مربع الخطأ في تقدير معلمة التمييز، كانت دالة إحصـائيًا في حالة اسـتخدام النموذج اللوجستي الثلاثي المعلمة مقارنة مع كل من النماذج الأخرى المستخدمة لمعالجة التخمين (الثنائي المعدل وعتبة التخمين والقائم على الصـعوبة)، حيث كانت قيمة مسـتوى الدلالة أقل من (٠,٠٥) في كل حالة؛ الأمر الذي يُشـير إلى أن النموذج اللوجسـتي الثلاثي المعلمة أدقّ في تقدير معلمة التمييز من النماذج الأخرى.

كما ويُلاحظ أن الفروق بين المتوسـطات الحسـابية للجذر التربيعي لمتوسـط مربع الخطأ في تقدير معلمة التمييز كانت دالة إحصـائيًا عند مسـتوى الدلالة (٠,٠٥) في حالة مقارنة النموذج الثنائي المعدل مع النموذج القائم على الصـعوبة؛ ممّا يُشـير إلى أن النموذج الثنائي المعدل أدق من هذا النموذج في تقدير معلمة التمييز، وكذلك تُشير النتائج إلى أنه عند مقارنة المتوسطات الحسابية للجذر التربيعي لمتوسـط مربع الخطأ لنموذج عتبة التخمين مع المتوسـطات الحسـابية نفسـها الخاصة بالنموذج القائم على الصعوبة كانت الفروق دالة إحصائيًا، الأمر الذي يُشير إلى أن نموذج عتبة التخمين أدق من النموذج القائم على الصعوبة في تقدير معلمة التمييز.

واستكمالًا للإجابة عن السؤال الأول، تمّ حساب المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة التمييز باختلاف كل من حجم العينة والنموذج المستخدم لمعالجة التخمين وطول الاختبار والجدول (٢١) يُبين نتائج هذا المؤشر الإحصائي.

جدول (٢١)

المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة التمييز والانحرافات المعيارية المناظرة لها في حالة النماذج الأربعة المختلفة المستخدمة لمعالجة التخمين وتبعًا لحجم عينة البيانات المولدة وعدد فقرات الاختبار.

حجم العينة	النموذج	طول الاختبار			
		٢٥ فقرة		٥٠ فقرة	
		متوسط التحيز	الانحراف المعياري	متوسط التحيز	الانحراف المعياري
٥٠٠ مفحوص	الثلاثي	٠,٠٣٤	٠,٠١٧	٠,٠١١	٠,٠١٢
	الثنائي المعدل	٠,٠٤٩	٠,٠٢٤	٠,٠٢١	٠,٠١٧
	عتبة التخمين	٠,٠٤١	٠,٠٣١	٠,٠٣٩	٠,٠٢٢
	القائم على الصعوبة	٠,٠٥٧	٠,٠٣٧	٠,٠٤٩	٠,٠٢٦
١٥٠٠ مفحوص	الثلاثي	٠٠٣٠	٠,٠١٧	٠,٠٠٤	٠,٠١٢
	الثنائي المعدل	٠,٠٣٥	٠,٠٢٤	٠,٠١١	٠,٠١٧
	عتبة التخمين	٠,٠٣٩	٠,٠٣١	٠,٠٣٤	٠,٠٢٢
	القائم على الصعوبة	٠,٠٤٣	٠,٠٣٧	٠,٠٣٠	٠,٠٢٦

ويُلاحظ من نتائج الجدول(٢١) انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التمييز (٠,٠٢٣) في حالة النموذج اللوجستي الثلاثي المعلمة، كما بلغ (٠,٠٣٥) في حالة النموذج الثنائي المعدل، يليه نموذج عتبة التخمين بقيمة بلغت(٠,٠٤١)، واخيرًا النموذج القائم على الصعوبة عتبة التخمين بقيمة بلغت(٠,٠٥٣).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ الانخفاض الظاهري لجميع قيم المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة التمييز، ففي حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,٠١٧)، يليه النموذج الثنائي المعدل بقيمة بلغت(٠,٠٢٣)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,٠٣٦)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٣٧).

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التمييز (٠,٠٣٢) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه نموذج عتبة التخمين بقيمة بلغت (٠,٠٤٠)، ثم النموذج الثنائي المعدل بقيمة بلغت(٠,٠٤٢)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت (٠,٠٥٠).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يُلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التمييز (٠,٠٠٨) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت(٠,٠١٦)، ثم نموذج عتبة التخمين بقيمة بلغت (٠,٠٣٧)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت (٠,٠٤٠).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التمييز (٠,٠٤١)، وانخفض ظاهريًا ليصل (٠,٠٢٥) في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التمييز (٠,٠٣٨) وفي حالة العينة المكونة من (١٥٠٠) مفحوص فقد انخفض ظاهريًا المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التمييز ليبلغ القيمة (٠,٠٢٨).

وللوقوف على أن التغييرات في قيم متوسطات مؤشر التحيز في تقدير معلمة التمييز تختلف تبعًا لمتغيرات الدراسة نموذج معالجة التخمين وطول الاختبار وحجم عينة المفحوصين المولدة لهم معلمة التمييز والتفاعلات بينها، تمّ استخدام نتائج تحليل التباين الثلاثي للقياسات المتكررة لتأثير هذه المتغيرات والتفاعل بينها على المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ لتقدير معلمة التمييز، والجدول رقم(٢٢) يبُين ملخص نتائج هذا التحليل.

جدول (٢٢)

ملخص نتائج تحليل التباين الثلاثي للقياسات المتكررة لأثر متغيرات الدراسة: النموذج المستخدم لمعالجة التخمين وطول الاختبار وحجم العينة والتفاعلات بينها على متوسطات مؤشر التحيز في تقدير معلمة التمييز.

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة الإحصائية
داخل الفقرات	النموذج.	٠,٦١٠	٣	٠,٢٠٣	٦,٠١١	٠,٠٠١*
	النموذج* عدد الفقرات.	٠,٠٦٣	٣	٠,٠٢١	٠,٦١١	٠,٦٠١
	النموذج*حجم العينة.	٠,٠٣٣	٣	٠,٠٥١	٠,٣٢٢	٠,٨١٤
	النموذج*عدد الفقرات*حجم العينة.	٠,٠٠٩	٣	٠,٠٠٣	٠,٠٨٨	٠,٩٦٧
	الخطأ	١٤,٨١٤	٤٣٨	٠,٠٣٤		
بين الفقرات	عدد الفقرات	٠,١٠٠	١	٠,١٠٠	٠,٩٠٨	٠,٣٤٢
	حجم العينة	٠,٠٢٩	١	٠,٠٢٩	٠,٢٥٨	٠,٦١٢
	عدد الفقرات × حجم العينة	٠,٠١٢	١	٠,٠١٢	٠,١١٢	٠,٧٣٨
	الخطأ	١٦,١٥٦	١٤٦	٠,١١١		

يُلاحظ من نتائج الجدول (٢٢) عدم وجود أثرٍ للتفاعلات الثلاثية أوالثنائية لمتغيرات الدراسة النموذج المستخدم لمعالجة التخمين وعدد الفقرات وعدد المفحوصين المولد لهم بيانات على متوسطات مؤشر التحيز في تقدير معلمة التمييز حيث كانت جميع قيم (ف) لها ليست دالة إحصائيًا، بينما يُلاحظ أن هناك أثرًا دالاً إحصائيًا يُعزى للنموذج المستخدم لمعالجة التخمين على متوسطات مؤشر التحيز في تقدير معلمة الصعوبة فقد كانت قيمة (ف) دالة إحصائيًا عند مستوى الدلالة (α=0.05).

وبالنسبة للتباين بين الافراد تُشير نتائج الجدول إلى عدم وجود أثرٍ للتفاعل الثنائي ولعدد الفقرات او لحجم عينة المفحوصين المولد لهم بيانات؛ فقد كانت جميع قيم (ف) لها ليست دالة إحصائيًا عند مستوى الدلالة(α=0.05).

ولمعرفة أي النماذج المستخدمة لمعالجة التخمين كان له تأثير دال إحصائيًا على متوسطات مؤشر التحيز في تقدير معلمة التمييز، تم اســـتخراج نتائج اختبار أقل فرق دال (LSD) والتي تبدو في الجدول الآتي:

جدول (٢٣)

نتائج اختبار أقل فرق دال (LSD) للمقارنات الثنائية بين متوسطات مؤشر التحيز في تقدير معلمة التمييز وباختلاف النموذج المستخدم لمعالجة التخمين.

النموذج (١)	النموذج (٢)	متوسط الفروق بين النموذجين (١- ٢)	مستوى الدلالة
الثلاثي المعلمة	الثنائي المعدل	-٠,٠٣٩.	٠,٠٧٦
	عتبة التخمين	-٠,٠٩٠	*٠,٠٠٠
	القائم على الصعوبة	-٠,٠٦٩	*٠,٠٠٢
الثنائي المعدل	الثلاثي المعلمة	٠,٠٣٩	٠,٠٧٦
	عتبة التخمين	-٠,٠٥١	*٠,٠٢٠
	القائم على الصعوبة	-٠,٠٣٠	٠,٢٢٣
عتبة التخمين	الثلاثي المعلمة	٠,٠٩٠	*٠,٠٠٠
	الثنائي المعدل	٠,٠٥١	*٠,٠٢٠
	القائم على الصعوبة	٠,٠٢١	٠,٣٦٨
القائم على الصعوبة	الثلاثي المعلمة	٠,٠٦٩	*٠,٠٠٢
	الثنائي المعدل	٠,٠٣٠	٠,٢٢٣
	عتبة التخمين	-٠,٠٢١	٠,٣٦٨

يُلاحظ من نتائج الجدول (٢٣) أن الفروق بين المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة التمييز للنماذج المستخدمة لمعالجة التخمين كانت دالة إحصائيًا في حالة استخدام النموذج اللوجستي الثلاثي المعلمة مقارنة مع النموذجين (عتبة التخمين والقائم على الصعوبة)؛ الأمر الذي يُشير إلى دقة النموذج اللوجستي الثلاثي المعلمة في تقدير معلمة التمييز مقارنة بهذين النموذجين، وعند مقارنة متوسطات مؤشر التحيز للنموذج الثنائي المعدل مع النماذج الأخرى فإن الفروق دالة إحصائيًا في حالة المقارنة مع نموذج عتبة التخمين فقد كانت مستوى الدلالة أقل من (٠,٠٥) الأمر الذي يُشير إلى دقة التقدير للنموذج الثنائي المعلمة المُعدل مقارنة مع نموذج عتبة التخمين، وعند مقارنة نموذج

عتبة التخمين مع النموذج القائم على الصعوبة فإن حجم متوسطات مؤشر التحيز أعلى في حالة نموذج عتبة التخمين إلا أنها غير دالة إحصائيًا.

ب- معلمة الصعوبة

الجدول (٢٤)

المتوسطات الحسابية لمؤشر الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة في حالة النماذج الأربعة المختلفة المستخدمة لمعالجة التخمين والانحرافات المعيارية المناظرة لها تبعًا لحجم عينة البيانات المولدة وعدد فقرات الاختبار.

حجم العينة	النموذج	طول الاختبار			
		٢٥ فقرة		٥٠ فقرة	
		المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
٥٠٠ مفحوص	الثلاثي	٠,٠٦٨	٠,٠١٣	٠,٠٤٣	٠,٠٠٩
	الثنائي المعدل	٠,٠٦١	٠,٠١٢	٠,٠٥٢	٠,٠٠٨
	عتبة التخمين	٠,١٥٣	٠,٠٣٩	٠,١٤٨	٠,٠٢٨
	القائم على الصعوبة	٠,٢٠٢	٠,٠٤٤	٠,١٥٥	٠,٠٣١
١٥٠٠ مفحوص	الثلاثي	٠,٠٢٩	٠,٠١٢	٠,٠١٦	٠,٠٠٩
	الثنائي المعدل	٠,٠٥٢	٠,٠١٣	٠,٠٤٦	٠,٠١٠
	عتبة التخمين	٠,١٤٦	٠,٠٣٩	٠,١١٢	٠,٠٢٨
	القائم على الصعوبة	٠,١٤٢	٠,٠٤٣	٠,١٣٢	٠,٠٣١

يُلاحظ من نتائج الجدول(٢٤) بالنسبة لمعلمة الصعوبة انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة (٠,٠٥٦) في حالة النموذج اللوجستي الثلاثي المعلمة، كما بلغ (٠,٠٥٧) في حالة النموذج الثنائي المعدل، يليه نموذج عتبة التخمين بقيمة بلغت(٠,١٥١). واخيرًا نموذج عتبة التخمين بقيمة بلغت(٠,١٧٨).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ الانخفاض الظاهري لجميع قيم المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة، ففي حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي (٠,٠٢٢)، يليه النموذج الثنائي المعدل بقيمة بلغت(٠,٠٤٧)، ثم نموذج عتبة التخمين بقيمة بلغت (٠,١٢٩)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,١٣٧)

ويُلاحظ من نتائج الجدول(٢٤) انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة(٠,٠٤٨) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت(٠,٠٥٦)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,١٤٩)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,١٧٢) .

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يّلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة (٠,٠٣٠) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت(٠,٠٤٧)، ثم نموذج عتبة التخمين بقيمة بلغت (٠,١٣٠)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,١٤٣). وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة (٠,١٠٧)، وانخفض ظاهريًا ليصل (٠,٠٨٠) في حالة الاختبار المكون من (٥٠) فقرة. وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة (٠,١١٠)، وفي حالة العينة المكونة من (١٥٠٠) مفحوص فقد انخفض ظاهريًا المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ ليبلغ القيمة (٠,٠٨٤).

وللوقوف على أن التغييرات في قيم متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة تختلف تبعًا لمتغيرات الدراسة: نموذج معالجة التخمين وطول الاختبار وحجم عينة المفحوصين المولد لهم معلمة الصعوبة والتفاعلات بينها، جرى استخدام نتائج تحليل التباين الثلاثي للقياسات المتكررة لتأثير هذه المتغيرات والتفاعل بينها على المتوسطات الحسابية للجذر التربيعي لمتوسط مربعات الخطأ في تقدير معلمة الصعوبة، والجدول (٢٥) يُبين ملخص نتائج هذا التحليل.

جدول (٢٥)

ملخص نتائج تحليل التباين الثلاثي للقياسات المتكررة لأثر متغيرات الدراسة: النموذج المستخدم لمعالجة التخمين وطول الاختبار وحجم العينة والتفاعلات بينها على متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة للفقرات.

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	ف	الدلالة الإحصائية
داخل الفقرات	النموذج	٢٦,٧٧٨	٣	٨,٩٢٦	١٦,٣١٧	٠,٠٠٠*
	النموذج* عدد الفقرات	١,٣٩٥	٣	٠,٤٦٥	٠,٨٥٠	٠,٤٦٧
	النموذج*حجم العينة	٠,٥٥٥	٣	٠,١٨٥	٠,٣٣٨	٠,٧٩٨
	النموذج*عدد الفقرات*حجم العينة	٠,١٧١	٣	٠,٠٥٧	٠,١٠٤	٠,٩٥٨
	الخطأ	٢٣٩,٥٩٦	٤٣٨	٠,٥٤٧		
بين الفقرات	عدد الفقرات	٦,٩٢٠	١	٦,٩٢٠	٢,٨٣٠	٠,٠٩٥
	حجم العينة	١,٣٨٧	١	١,٣٨٧	٠,٥٦٧	٠,٤٥٣
	عدد الفقرات × حجم العينة	٠,٦٢٢	١	٠,٦٢٢	٠,٢٥٤	٠,٦١٦
	الخطأ	٣٥٦,٩٨٠	١٤٦	٢,٤٤٥		

يُلاحظ من الجدول (٢٥) عدم وجود أثرٍ للتفاعلات الثلاثية والثنائية لمتغيرات الدراسة النموذج المستخدم لمعالجة التخمين وعدد الفقرات وعدد المفحوصين المولد لهم بيانات على متوسطات الجذر التربيعي لمتوسط مربع الخطأ لتقدير معلمة الصعوبة حيث كانت جميع قيم (ف) لها ليست دالة إحصائيًا، بينما يُلاحظ أن هناك أثرًا دالاً إحصائيًا يُعزى للنموذج المستخدم لمعالجة التخمين على المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة فقد كانت قيمة (ف) دالة إحصائيًا عند مستوى الدلالة (α=0.05).

وبالنسبة للتباين بين الافراد تُشير نتائج الجدول إلى عدم وجود أثرٍ للتفاعل الثنائي أو لعدد الفقرات او لحجم عينة المفحوصين المولد لهم بيانات؛ فقد كانت جميع قيم (ف) لها ليست دالة إحصائيًا عند مستوى الدلالة(α=0.05). ولمعرفة أي النماذج المستخدمة لمعالجة التخمين كان له تأثير دالّ إحصائيًا على متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة، تم إجراء اختبار أقل فرق دال (LSD) والذي تبدو نتائجه فيما يلي.

جدول (٢٦)

نتائج اختبار أقل فرق دال (LSD) بين متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة تبعًا للنموذج المستخدم.

النموذج (١)	النموذج (٢)	متوسط الفروق بين النموذجين (١- ٢)	الدلالة الإحصائية
الثلاثي المعلمة	الثنائي المعدل	-٠,٠١٣	*٠,٠٤٩
	عتبة التخمين	-٠,١٠١	*٠,٠٠٠
	القائم على الصعوبة	-٠,١١٩	*٠,٠٠٠
الثنائي المعدل	الثلاثي المعلمة	٠,٠١٣	*٠,٠٤٩
	عتبة التخمين	- ٠,٠٨٨	*٠,٠٠٠
	القائم على الصعوبة	-٠,١٠٦	*٠,٠٠٠
عتبة التخمين	الثلاثي المعلمة	٠,١٠١	*٠,٠٠٠
	الثنائي المعدل	٠,٠٨٨	*٠,٠٠٠
	القائم على الصعوبة	٠,٠١٨	٠,١٤٧
القائم على الصعوبة	الثلاثي المعلمة	٠,١١٩	*٠,٠٠٠
	الثنائي المعدل	٠,١٠٦	*٠,٠٠٠
	عتبة التخمين	-٠,٠١٨	٠,١٤٧

يُلاحظ من نتائج الجدول (٢٦) وعند مقارنة المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة للنموذج اللوجستي الثلاثي المعلمة بالنماذج الأخرى فإن الفروق دالة إحصائيًا عند مستوى الدلالة (α=0.05) عند كل مقارنة، الأمر الذي يُشير إلى أن النموذج اللوجستي الثلاثي المعلمة أدق في تقدير معلمة الصعوبة مقارنة مع تلك النماذج.

وعند مقارنة متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة للنموذج الثنائي المعدل مع كل من النماذج الأخرى المستخدمة لمعالجة التخمين، كانت الفروق دالة إحصائيًا، الأمر الذي يُشير إلى دقة النموذج الثنائي المعدل في تقدير معلمة الصعوبة، وعند مقارنة متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة الصعوبة للنموذج القائم على الصعوبة مع نموذج عتبة التخمين لم تكن الفروق دالة إحصائيًا.

واستكمالاً للإجابة عن السؤال الثاني، فقد جرى حساب المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة الصعوبة باختلاف كل من حجم العينة والنموذج اللوجستي وطول الاختبار، والجدول رقم (٢٧) يبين نتائج هذا المؤشر الإحصائي.

جدول (٢٧)

المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة الصعوبة والانحرافات المعيارية المناظرة لها في حالة النماذج الأربعة المختلفة المستخدمة لمعالجة التخمين تبعًا لحجم عينة البيانات المولدة وعدد فقرات الاختبار.

حجم العينة	النموذج	طول الاختبار			
		٢٥ فقرة		٥٠ فقرة	
		المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
٥٠٠ مفحوص	الثلاثي	٠,٠٣٢	٠,٠٣٨	٠,٠٣٧	٠,٠٢٧
	الثنائي المعدل	٠,٠٥٧	٠,٠٤٤	٠,٠٤٩	٠,٠٣١
	عتبة التخمين	٠,٠٦٩	٠,٠٧٣	٠,٠٦٧	٠,٠٥٢
	القائم على الصعوبة	٠,٠٧٣	٠,٠٧٦	٠,٠٧٠	٠,٠٥٤
١٥٠٠ مفحوص	الثلاثي	٠,٠١٢	٠,٠٣٢	٠,٠٠٣	٠,٠٢٤
	الثنائي المعدل	٠,٠٤٦	٠,٠٤١	٠,٠٤٦	٠,٠٣١
	عتبة التخمين	٠,٠٦٥	٠,٠٦١	٠,٠٦٠	٠,٠٥٢
	القائم على الصعوبة	٠,٠٦٨	٠,٠٦٨	٠,٠٦٥	٠,٠٥٤

يُلاحظ من نتائج الجدول(٢٧) انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة الصعوبة (٠,٠٣٤) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعدل بقيمة بلغت (٠,٠٥٣)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,٠٦٨)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٧٢).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ الانخفاض الظاهري لجميع قيم المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة الصعوبة، ففي حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,٠٠٨)، يليه النموذج الثنائي المُعدل بقيمة بلغت(٠,٠٤٦)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,٠٦٣)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٦٦).

كما يُلاحظ من نتائج الجدول انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة الصعوبة (٠,٠٢٢) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت(٠,٠٥٢)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,٠٦٧)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٧١).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يلاحظ أن جميع قيم متوسطات مؤشر التحيز في تقدير معلمة الصعوبة قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة الصعوبة (٠,٠٢٠) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعلمة المعدل بقيمة بلغت(٠,٠٤٨)، ثم نموذج عتبة التخمين بقيمة بلغت(٠,٠٦٣)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٦٨).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة الصعوبة (٠,٠٥٣)، وانخفض ظاهريًا ليصل (٠,٠٤٩) في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة الصعوبة (٠,٠٥٧)، وفي حال العينة المكونة من (١٥٠٠) مفحوص فقد انخفض ظاهريًا المتوسط الحسابي لمؤشر التحيز في تقدير معلمة الصعوبة ليبلغ القيمة (٠,٠٤٦).

وللوقوف على أن التغييرات في قيم متوسطات مؤشر التحيز في تقدير معلمة الصعوبة تختلف تبعًا لمتغيرات الدراسة: نموذج معالجة التخمين وطول الاختبار وحجم عينة المفحوصين المولدة لهم معلمة الصعوبة والتفاعلات بينها جرى استخدام نتائج تحليل التباين الثلاثي للقياسات المتكررة لتأثير هذه المتغيرات والتفاعل بينها على المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة الصعوبة والجدول رقم (٢٨) يبين ملخص نتائج هذا التحليل.

جدول (٢٨)

ملخص نتائج تحليل التباين الثلاثي للقياسات المتكررة لأثر متغيرات الدراسة: النموذج المستخدم لمعالجة التخمين وطول الاختبار وحجم العينة والتفاعلات بينها على متوسطات مؤشر التحيز في تقدير معلمة الصعوبة للفقرات.

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	ف	الدلالة الإحصائية
داخل الفقرات	النموذج	٢,٥٦٧	٣	٠,٨٥٦	٣٥,٤٨٥	٠,٠٠٠*
	النموذج* عدد الفقرات	٠,٠٤٩	٣	٠,٠١٦	٠,٦٨٠	٠,٥٦٣
	النموذج*حجم العينة	٠,٠٧٦	٣	٠,٠٢٥	١,٠٥١	٠,٣٧٠
	النموذج*عدد الفقرات*حجم العينة	٠,٠٨١٢٤	٣	٠,٠٠٨	٠,٣٢٨	٠,٨٠٥
	الخطأ	١٠,٥٦٣	٤٣٨	٠,٠٢٤		
بين الفقرات	عدد الفقرات	٠,٠٩٦	١	٠,٠٩٦	١,٤٦٤	٠,٢٢٨
	حجم العينة	٠,١٨١	١	٠,١٨١	٢,٧٦١	٠,٠٩٩
	عدد الفقرات × حجم العينة	٠,٠٢٠	١	٠,٠٢٠	٠,٣٠١	٠,٥٨٤
	الخطأ	٩,٥٦٥	١٤٦		٠,٠٦٦	

يُلاحظ من نتائج الجدول (٢٨) عدم وجود أثرٍ ذي دلالة إحصـــائية للتفاعلات الثلاثية أو الثنائية لمتغيرات الدراســة وهي النموذج المســتخدم لمعالجة التخمين وعدد الفقرات وعدد المفحوصــين المولد لهم بيانات على متوسطات مؤشر التحيز في تقدير معلمة الصعوبة، فقد كانت قيم (ف) لها

ليست دالة إحصائيًا، بينما يُلاحظ أن هناك أثرًا يُعزى للنموذج المستخدم لمعالجة التخمين فقد بلغت قيمة ف (٣٥,٤٨٥)، وهي دالة إحصائيًا عند مستوى الدلالة (α=0.05).

وبالنسبة للتباين بين الافراد تُشير نتائج الجدول إلى عدم وجود أثرٍ للتفاعل الثنائي او لحجم العينة أو لعدد الفقرات او لعدد المفحوصين المولد لهم بيانات؛ فقد كانت جميع قيم (ف) لها ليست دالة إحصائيًا عند مستوى الدلالة (α=0.05).

ولمعرفة أي النماذج المستخدمة لمعالجة التخمين كان له تأثير دال إحصائيًا على قيم متوسطات مؤشر التحيز في تقدير معلمة الصعوبة، تم استخراج نتائج اختبار اقل فرق دال (LSD) Least Significant Differences والتي تبدو في الجدول الآتي:

جدول (٢٩)

نتائج اختبار أقل فرق دال (LSD) للمقارنات الثنائية بين متوسطات مؤشر التحيز في تقدير معلمة الصعوبة تبعًا للنموذج المستخدم.

النموذج (١)	النموذج (٢)	متوسط الفروق بين النموذجين (١ - ٢)	مستوى الدلالة
الثلاثي المعلمة	الثنائي المعدل	-٠,٠٤٣	٠,٠٠٢*
	عتبة التخمين	-٠,١٥٥	٠,٠٠٠*
	القائم على الصعوبة	-٠,١٥٨	٠,٠٠٠*
الثنائي المعدل	الثلاثي المعلمة	٠,٠٤٣	٠,٠٠٢*
	عتبة التخمين	-٠,١١٣	٠,٠٠٠*
	القائم على الصعوبة	-٠,١١٦	٠,٠٠٠*
عتبة التخمين	الثلاثي المعلمة	٠,١٥٥	٠,٠٠٠*
	الثنائي المعدل	٠,١١٣	٠,٠٠٠*
	القائم على الصعوبة	-٠,٠٠٣	٠,٨٩٤
القائم على الصعوبة	الثلاثي المعلمة	٠,١٥٨	٠,٠٠٠*
	تثبيت معلمة التخمين	٠,١١٦	٠,٠٠٠*
	عتبة التخمين	٠,٠٠٣	٠,٨٩٤

يُلاحظ من نتائج الجدول (٢٩) أن الفروق بين المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة الصعوبة لنماذج معالجة التخمين كانت دالة إحصائيًا في حالة مقارنة النموذج اللوجستي الثلاثي المعلمة مع كل من النموذج الثنائي المعدل و نموذج عتبة التخمين والنموذج القائم على الصعوبة حيث كانت قيمة مستوى الدلالة أقل من (٠,٠٥) عند كل مقارنة؛ الأمر الذي يُشير إلى دقة التقدير

للنموذج اللوجستي الثلاثي المعلمة مقارنة بهذه النماذج المستخدمة لمعالجة التخمين، وعند مقارنة متوسطات مؤشر التحيز للنموذج الثنائي المُعدل مع النماذج الأخرى، وجُد إن الفروق دالة إحصائيًا عند المقارنة مع كل من النموذج القائم على الصعوبة ونموذج عتبة التخمين، الأمر الذي يُشير إلى دقة التقديرات للنموذج الثنائي المعدل في تقدير معلمة الصعوبة للمفحوصين مقارنة بهذين النموذجين، وعند مقارنة نموذج عتبة التخمين مع النموذج القائم على الصعوبة فإن حجم المتوسطات الحسابية لمؤشر التحيز أعلى في حالة النموذج القائم على الصعوبة إلا أنها غير دالة إحصائيًا.

ج-معلمة التخمين:

الجدول (٣٠)

المتوسطات الحسابية والانحرافات المعيارية المناظرة لها لمؤشر الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين في حالة النماذج الأربعة المختلفة المستخدمة لمعالجة التخمين تبعًا لحجم عينة البيانات المولدة وعدد فقرات الاختبار.

حجم العينة	النموذج	طول الاختبار			
		٢٥ فقرة		٥٠ فقرة	
		المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
٥٠٠ مفحوص	الثلاثي	٠,٠٧٧	٠,٠١٠	٠,٠٥٨	٠,٠١٢
	الثنائي المعدل	*	*	*	*
	عتبة التخمين	٠,٠٩٨	٠,٠١١	٠,٠٨٦	٠,٠٠٨
	القائم على الصعوبة	٠,١٢٤	٠,٠١٥	٠,٠٧٨	٠,٠١٢
١٥٠٠ مفحوص	الثلاثي	٠,٠٥٧	٠,٠٠٧	٠,٠٣١	٠,٠٠٦
	الثنائي المعدل	*	*	*	*
	عتبة التخمين	٠,٠٨٦	٠,٠٠٨	٠,٠٨١	٠,٠٠٧
	القائم على الصعوبة	٠,٠٧٧	٠,٠١٠	٠,٠٦٨	٠,٠١١

* تم تثبيت معلمة التخمين عند القيمة ٠,٢٠

يُلاحظ من نتائج الجدول انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين(٠,٠٦٧)، في حالة النموذج اللوجستي الثلاثي المعلمة، يليه نموذج عتبة التخمين بقيمة بلغت(٠,٠٩٢)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,١٠١).

وعند الانتقال إلى العينة المكونة من (١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ الانخفاض الظاهري لجميع قيم الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين، ففي حالة النموذج اللوجستي الثلاثي المعلمة بلغ المتوسط الحسابي(٠,٠٤٤)، يليه النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٧٣)، واخيرًا نموذج عتبة التخمين بقيمة بلغت(٠,٠٨٤)

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين (٠,٠٦٧) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه نموذج عتبة التخمين بقيمة بلغت(٠,١٠٤)، واخيرًا النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٩٠).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يُلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين (٠,٠٤٤) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٧٣)، واخيرًا نموذج عتبة التخمين بقيمة بلغت (٠,٠٨٣).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين (٠,٠٨٦)، وانخفض ظاهريًا ليصل (٠,٠٦٧) في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين (٠,٠٨٧).

وفي حالة العينة المكونة من (١٥٠٠) مفحوص انخفض ظاهريًا المتوسط الحسابي للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين ليبلغ القيمة (٠,٠٦٦).

وللوقوف على أن التغييرات في قيم متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين تختلف تبعًا لمتغيرات الدراسة: نموذج معالجة التخمين وطول الاختبار وحجم عينة المفحوصين المولد لهم معلمة التخمين والتفاعلات بينها، تمّ استخدام نتائج تحليل التباين الثلاثي

للقياسات المتكررة لتأثير هذه المتغيرات والتفاعل بينها على المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين والجدول (٣١) يبين ملخص نتائج هذا التحليل.

جدول (٣١)

ملخص نتائج تحليل التباين الثلاثي للقياسات المتكررة لأثر متغيرات الدراسة: النموذج المستخدم لمعالجة التخمين وطول الاختبار وحجم العينة والتفاعلات بينها على المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين.

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	ف	الدلالة الإحصائية
داخل الفقرات	النموذج	٠,٠١٤	٢	٠,٠٠٧	١١,٤٦٠	٠,٠٠٠*
	النموذج* عدد الفقرات	٠,٠٠٣	٢	٠,٠٠١	٢,٠٨٩	٠,١٢٦
	النموذج*حجم العينة	٠,٠٠٦	٢	٠,٠٠٣	٤,٤٢٧	٠,٠٢٣
	النموذج*عدد الفقرات*حجم العينة	٠,٠٠٣	٢	٠,٠٠٢	٢,٤٢٦	٠,٠٩٠
	الخطأ	٠,١٨٢	٢٩٢	٠,٠٠١		
بين الفقرات	عدد الفقرات	٠,٠٠٣	١	٠,٠٠٣	٢,٥٢٩	٠,١١٤
	حجم العينة	٠,٠٠٠٣	١	٠,٠٠٣	٠,٠٢٨	٠,٨٦٧
	عدد الفقرات × حجم العينة	٠,٠٠٠	١	٠,٠٠٠	٠,٤٠٤	٠,٥٢٦
	الخطأ	٠,١٦٢	١٤٦	٠,٠٠١		

يُلاحظ من الجدول (٣١)، عدم وجود أثرٍ للتفاعل الثلاثي بين النموذج المستخدم لمعالجة التخمين وعدد الفقرات وحجم العينة أو للتفاعل الثنائي بين النموذج المستخدم وعدد الفقرات على متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين حيث كانت جميع قيم (ف) لها ليست

دالة إحصائيًا عند مستوى الدلالة (α=0.05)، بينما يُلاحظ وجود أثر للنموذج المستخدم لمعالجة التخمين، فقد كانت قيمة (ف) دالة إحصائيًا عند مستوى الدلالة (α=0.05).

وبالنسبة للتباين بين الأفراد تُشير نتائج الجدول إلى عدم وجود أثرٍ للتفاعل الثنائي أو لعدد الفقرات أو لعدد المفحوصين المولد لهم بيانات؛ فقد كانت جميع قيم (ف) لها ليست دالة إحصائيًا عند مستوى الدلالة (α=0.05).

ولمعرفة أي النماذج المستخدمة لمعالجة التخمين كان له تأثير دال إحصائيًا على متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين تم استخراج نتائج اختبار أقل فرق دال Least Significant Differences (LSD) والتي تبدو في الجدول (٣٢).

جدول (٣٢)

نتائج اختبار أقل فرق دال (LSD) للمقارنات الثنائية بين متوسطات الجذر التربيعي لمتوسط مربع الخطأ في تقدير معلمة التخمين.

النموذج (١)	النموذج (٢)	متوسط الفروق بين النموذجين (١- ٢)	الدلالة الإحصائية
الثلاثي المعلمة	عتبة التخمين	-٠,٠٣٢	*٠,٠٠٠
	القائم على الصعوبة	-٠,٠٣٤	*٠,٠٠٠
عتبة التخمين	الثلاثي المعلمة	٠,٠٣٢	*٠,٠٠٠
	القائم على الصعوبة	-٠,٠٠٢	٠,٧٩٧
القائم على الصعوبة	الثلاثي المعلمة	٠,٠٣٤	*٠,٠٠٠
	عتبة التخمين	٠,٠٠٢	٠,٧٩٧

يُلاحظ من الجدول (٣٢) وجود فروق ذات دلالة إحصـــائية بين المتوســـطات الحســـابية للجذر التربيعي لمتوســـط مربع الخطأ في تقدير معلمة التخمين تبعًا للنموذج المســـتخدم لمعالجة التخمين في حالة النموذج اللوجســـتي الثلاثي المعلمة عند مقارنته مع كل من نموذج عتبة التخمين والنموذج القائم على الصــعوبة، إذ كانت قيمة الدلالة أقل من (٠,٠٥) عند كل مقارنة، الأمر الذي يُشـــير إلى أن دقة النموذج اللوجســـتي الثلاثي المعلمة في تقدير معلمة التخمين، بينما لم تكن الفروق دالة إحصـائيًا بين المتوسـطات الحسـابية في حالة مقارنة نموذج عتبة التخمين مع النموذج القائم على الصعوبة.

واستكمالاً للإجابة عن السؤال الثاني، تمّ حساب المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة التخمين باختلاف كل من حجم العينة والنموذج المستخدم لمعالجة التخمين وطول الاختبار والجدول (٣٣) يُبين نتائج هذا المؤشر الإحصائي.

جدول (٣٣)

المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة التخمين والانحرافات المعيارية المناظرة لها في حالة النماذج الأربعة المختلفة المستخدمة لمعالجة التخمين تبعًا لحجم عينة البيانات المولدة وعدد فقرات الاختبار.

حجم العينة	النموذج	طول الاختبار			
		٢٥ فقرة		٥٠ فقرة	
		المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
٥٠٠ مفحوص	الثلاثي	٠,٠١٠٠	٠,٠١٥	٠,٠١٩٠	٠,٠١٠٠
	الثنائي المعدل	*	*	*	*
	عتبة التخمين	٠,٠٥٢٠	٠,٠٢٦	٠,٠٣٧٠	٠,٠١٤٠
	القائم على الصعوبة	٠,٠٥٨٠	٠,٠٢٢	٠,٠١١٠	٠,٠١٦٠
١٥٠٠ مفحوص	الثلاثي	٠,٠٠١١	٠,٠١٥	٠,٠٠١٠	٠,٠١٥٠
	الثنائي المعدل	*	*	*	*
	عتبة التخمين	٠,٠٢١	٠,٠٢٠	٠,٠١٦٠	٠,٠١٤٠
	القائم على الصعوبة	٠,٠١٥	٠,٠٠٢٢	٠,٠٠٦٠	٠,٠١٦٠

* تم تثبيت معلمة التخمين عند القيمة ٠,٢٠

ويُلاحظ من نتائج الجدول انه في حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار، بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التخمين (٠,٠١٥) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج القائم على الصعوبة بقيمة بلغت(٠,٠٣٥)، واخيرًا نموذج عتبة التخمين بقيمة بلغت(٠,٠٤٥).

وعند الانتقال إلى العينة المكونة من(١٥٠٠) مفحوص وبغض النظر عن طول الاختبار، يُلاحظ الانخفاض الظاهري لجميع قيم المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة التخمين، ففي حالة النموذج اللوجستي الثلاثي المعلمة ونموذج عتبة التخمين بلغ المتوسط الحسابي(٠,٠١٠)، ثم النموذج القائم على الصعوبة بقيمة بلغت(٠,٠١١).

كما يُلاحظ انه في حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين، بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التخمين (٠,٠٠٦) في حالة النموذج اللوجستي الثلاثي المعلمة، يليه النموذج القائم على الصعوبة ونموذج عتبة التخمين بقيمة بلغت(٠,٠٣٧).

وفي حالة الاختبار المكون من (٥٠) فقرة وبغض النظر عن حجم عينة المفحوصين يُلاحظ أن جميع القيم قد انخفضت ظاهريًا، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التخمين (٠,٠١٠) في حالة النموذج اللوجستي الثلاثي المعلمة والنموذج القائم على الصعوبة، واخيرًا نموذج عتبة التخمين بقيمة بلغت(٠,٠٢٧).

وفي حالة الاختبار المكون من (٢٥) فقرة وبغض النظر عن حجم عينة المفحوصين وعن النموذج المستخدم لمعالجة التخمين، فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التخمين (٠,٠٢٧)، وانخفض ظاهريًا ليصل (٠,٠٢٠) في حالة الاختبار المكون من (٥٠) فقرة.

وفي حالة العينة المكونة من (٥٠٠) مفحوص وبغض النظر عن طول الاختبار وعن النموذج المستخدم لمعالجة التخمين فقد بلغ المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التخمين (٠,٠٣٢) وفي حال العينة المكونة من (١٥٠٠) مفحوص انخفض ظاهريًا المتوسط الحسابي لمؤشر التحيز في تقدير معلمة التخمين ليبلغ القيمة (٠,٠١٢).

وللوقوف على أن التغييرات في قيم متوسطات مؤشر التحيز في تقدير معلمة التخمين تختلف تبعًا لمتغيرات الدراسة: نموذج معالجة التخمين وطول الاختبار وحجم عينة المفحوصين المُولد لهم معلمة التخمين والتفاعلات بينها، تمّ استخدام نتائج تحليل التباين الثلاثي للقياسات المتكررة لتأثير هذه المتغيرات والتفاعل بينها على متوسطات مؤشر التحيز، والجدول (٣٤) يُبين ملخص نتائج هذا التحليل.

جدول (٣٤)

ملخص نتائج تحليل التباين الثلاثي للقياسات المتكررة لأثر متغيرات الدراسة: النموذج المُستخدم لمعالجة التخمين وطول الاختبار وحجم عينة المفحوصين والتفاعلات بينها على متوسطات مؤشر التحيز في تقدير معلمة التخمين.

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	ف	الدلالة الإحصائية
داخل الفقرات	النموذج	٠,١٨٩	٢	٠,٠٩٤	٧,٣٥١	٠,٠٠١*
	النموذج* عدد الفقرات	٠,٠١١	٢	٠,٠٠٥	٠,٤٠٨	٠,٦٦٥
	النموذج*حجم العينة	٠,٠٠١	٢	٠,٠٠٠	٠,٠٢٤	٠,٩٧٦
	النموذج*عدد الفقرات*حجم العينة	٠,٠٠٦	٢	٠,٠٠٣	٠,٢٣٠	٠,٧٩٥
	الخطأ	٣,٧٥٣	٢٩٢	٠,٠١٣		
بين الفقرات	عدد الفقرات	٠,٠٢٩	١	٠,٠٢٩	١,٧٧٣	٠,١٨٥
	حجم العينة	٠,٠٣١	١	٠,٠٣١	١,٨٥٨	٠,١٧٥
	عدد الفقرات × حجم العينة	٠,٠٠٣	١	٠,٠٠٣	٠,١٦١	٠,٦٨٩
	الخطأ	٢,٤٠٤	١٤٦	٠,٠١٦		

يُلاحظ من الجدول (٣٤) عدم وجود أثرٍ للتفاعل الثلاثي بين النموذج المستخدم لمعالجة التخمين وعدد الفقرات وحجم العينة أو للتفاعل الثنائي بين النموذج المستخدم وعدد الفقرات على متوسطات مؤشر التحيز في تقدير معلمة التخمين حيث كانت جميع قيم (ف) لها ليست دالة إحصائيًا عند مستوى الدلالة (α=0.05)، بينما يُلاحظ وجود أثرٍ للنموذج المستخدم لمعالجة التخمين، فقد كانت قيمة (ف) دالة إحصائيًا.

وبالنسبة للتباين بين الأفراد تُشير نتائج الجدول إلى عدم وجود أثرٍ للتفاعل الثنائي ولعدد الفقرات ولعدد المفحوصين المُولد لهم بيانات؛ فقد كانت جميع قيم (ف) لها ليست دالة إحصائيًا عند مستوى الدلالة (α=0.05).

ولمعرفة أي النماذج المستخدمة لمعالجة التخمين كان له تأثير دال إحصائيًا على متوسطات مؤشر التحيز في تقدير معلمة التخمين، تم استـــخراج نتائج اختبار أقل فرق دال (LSD) والتي تبدو في الجدول الآتي:

جدول (٣٥)

نتائج اختبار أقل فرق دال (LSD) للمقارنات الثنائية بين متوسطات مؤشر التحيز في تقدير معلمة التخمين، تبعًا للنموذج المستخدم لمعالجة التخمين.

النموذج (١)	النموذج (٢)	متوسط الفروق بين النموذجين (١- ٢)	مستوى الدلالة
الثلاثي المعلمة	عتبة التخمين	-٠,٠٣٤	*٠,٠٠٠
	القائم على الصعوبة	-٠,٠١٨	٠,٠٥٧
عتبة التخمين	الثلاثي المعلمة	٠,٠٣٤	*٠,٠٠٠
	القائم على الصعوبة	٠,٠١٧	٠,٠٥٣
القائم على الصعوبة	الثلاثي المعلمة	٠,٠١٨	٠,٠٥٧
	عتبة التخمين	-٠,٠١٧	٠,٠٥٣

يُلاحظ من نتائج الجدول (٣٥) أن الفروق بين المتوسطات الحسابية لمؤشر التحيز في تقدير معلمة التخمين كانت دالة إحصائيًا في حالة استخدام النموذج اللوجستي الثلاثي المعلمة مع نموذج عتبة التخمين حيث كانت قيمة (ف) دالة احصائيا عند مستوى الدلالة (α=0.05).

وعند المقارنة مع النموذج القائم على الصعوبة كانت قيم متوسطات التحيز في تقدير معلمة التخمين أعلى منها في حالة النموذج القائم القائم على الصعوبة إلا انها لم تكن دالة احصائيًا، الأمرالذي يُشير إلى دقة النموذج اللوجستي الثلاثي المعلمة مقارنة بهذه النماذج المستخدمة لمعالجة التخمين.

وعند مقارنة متوسطات مؤشر التحيز في تقدير معلمة التخمين لنموذج عتبة التخمين مع النموذج القائم على الصعوبة فإن حجم متوسطات مؤشر التحيز أعلى منها في حالة النموذج القائم على الصعوبة إلا أنها غير دالة إحصائيًا عند مستوى الدلالة (α=0.05).

النتائج المتعلقة بالإجابة عن سؤال الدراسة الثالث:

ما أثر النموذج المستخدم في الكفاءة النسبية لتقدير معالم الفقرة(الصعوبة والتمييز والتخمين) ومعلمة القدرة في ضوء اختلاف عدد الفقرات، حجم العينة؟

للإجابة عن هذا السؤال، تم حساب التباين للقيم المُقدرة وفق كل نموذج في المواقف المختلفة للمتغيرات، تم قسمة تباين القيم المقدرة لمعالم الفقرات (معلمة الصعوبة ومعلمة التمييز ومعلمة التخمين) لكل نموذج على تباين القيم المقدرة لنفس المعالم باستخدام النماذج الأخرى.

أولاً : إيجاد الكفاءة النسبية لمعلمة التمييز.

قيم الكفاءة النسبية لمعلمة التمييز وفقًا للنموذج المستخدم وبإختلاف طول الاختبار وحجم العينة.

جدول رقم (٣٦)

قيم الكفاءة النسبية لمعلمة التمييز وفقاً لكل نموذج من نماذج معالجة التخمين وباختلاف عدد الفقرات وحجم عينة البيانات المولدة.

طول الاختبار	النموذج	الخطأ المعياري للتقدير	
		٥٠٠ مفحوص	١٥٠٠ مفحوص
٢٥ فقرة	الثلاثي	٠,٠٩٥	٠,١٢٠
	الثنائي المعدل	٠,١١٦	٠,١١٨
	عتبة التخمين	٠,١٦٨	٠,١٤٣
	القائم على الصعوبة	٠,٣٠٧	٠,١٤٦
	الكفاءة النسبية لـ الثلاثي على الثنائي المعدل	١,٤٩	٠,٩٧
	الكفاءة النسبية لـ الثلاثي على عينة التخمين	٣,١٣	١,٤٢
	الكفاءة النسبية لـ الثلاثي على القائم على الصعوبة	١٠,٤٤	١,٤٨
	الكفاءة النسبية لـ الثنائي المعدل على عتبة التخمين	٢,١٠	١,٤٧
	الكفاءة النسبية لـ الثنائي المعدل على القائم على الصعوبة	٧,٠٠	١,٥٣
	الكفاءة النسبية لـ عتبة التخمين على القائم على الصعوبة	٣,٣٤	١,٠٤
٥٠ فقرة	الثلاثي	٠,٠٥٩	٠,٠٧٣
	الثنائي المعدل	٠,١٢٠	٠,١١٥
	عتبة التخمين	٠,١٦٢	٠,١٣٧
	القائم على الصعوبة	٠,١٥٧	٠,١٣٣
	الكفاءة النسبية لـ الثلاثي على الثنائي المعدل	٤,١٤	٢,٤٨
	الكفاءة النسبية لـ الثلاثي على عتبة التخمين	٧,٥٤	٣,٥٢
	الكفاءة النسبية لـ الثلاثي على القائم على الصعوبة	٧,٠٨	٣,٣٢
	الكفاءة النسبية لـ الثنائي المعدل على عتبة التخمين	١,٨٢	١,٤٢
	الكفاءة النسبية لـ الثنائي المعدل على القائم على الصعوبة	١,٧١	١,٣٤
	الكفاءة النسبية لـ عتبة التخمين على القائم على الصعوبة	٠,٩٤	٠,٩٤

يُلاحظ من الجدول (٣٦) ان النتائج الخاصة به قد جاءت على النحو الآتي:

اولاً: في حالة الاختبار المُكون من (٢٥) فقرة والمُولد عليه (٥٠٠) استجابة مفحوص:

أن قيم الكفاءة النسبية لمعلمة التمييز للنموذج الثلاثي المعلمة على النماذج الأخرى قد كانت ظاهريًا لصالح النموذج اللوجستي الثلاثي المعلمة، وكانت قيم الكفاءة النسبية لمعلمة التمييز لصالح النموذج الثنائي المعدل عند مقارنته مع كل من: نموذج عتبة التخمين والنموذج القائم على الصعوبة، وكانت قيم الكفاءة النسبية لمعلمة التمييز لصالح نموذج عتبة التخمين عند المقارنة مع النموذج القائم على الصعوبة.

ثانياً: في حالة الاختبار المُكون من (٥٠) فقرة والمُولد عليه (٥٠٠) استجابة لمفحوص:

ان قيم الكفاءة النسبية لمعلمة التميز قد كانت لصالح النموذج اللوجستي الثلاثي المعلمة عند مقارنته مع النماذج الأخرى، وعند مقارنة قيم الكفاءة النسبية لمعلمة التمييز للنموذج الثنائي المعدل مع كل من نموذج عتبة التخمين والنموذج القائم على الصعوبة كانت المقارنة لصالح النموذج الثنائي المعدل، وعند مقارنة نموذج عتبة التخمين مع النموذج القائم على الصعوبة، كانت قيمة الكفاءة النسبية لصالح النموذج القائم على الصعوبة.

ثالثاً: في حالة الاختبار المُكون من (٢٥) فقرة والمُولد عليه (١٥٠٠) استجابة لمفحوص:

كانت قيم الكفاءة النسبية لمعلمة التمييز لصالح النموذج اللوجستي الثلاثي المعلمة المعدل عند المقارنة مع النماذج الأخرى، ولصالح النموذج الثنائي المعدل عند المقارنة معه، وعند مقارنة النموذج الثنائي المعدل مع نموذج عتبة التخمين والنموذج القائم على الصعوبة كانت قيم الكفاءة النسبية لصالح النموذج الثنائي المعدل، واخيرًا كانت قيم الكفاءة النسبية لمعلمة التمييز لصالح نموذج عتبة التخمين عند المقارنة مع النموذج القائم على الصعوبة.

رابعًا: في حالة الاختبار المُكون من (٥٠) فقرة والمُولد عليه (١٥٠٠) استجابة لمفحوص:

كانت قيم الكفاءة النسبية لمعلمة التمييز لصالح النموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النماذج الأخرى المستخدمة لمعالجة التخمين، وعند مقارنة النموذج الثنائي المعدل مع نموذج عتبة التخمين والنموذج القائم على الصعوبة كانت قيم الكفاءة النسبية لمعلمة التمييز لصالح النموذج الثنائي المعدل، واخيرًا كانت قيم الكفاءة النسبية لصالح النموذج القائم على الصعوبة عند مقارنته مع نموذج عتبة التخمين.

ثانيا : إيجاد الكفاءة النسبية لمعلمة الصعوبة.

جدول (٣٧)

قيم الكفاءة النسبية لمعلمة الصعوبة وفقاً لكل نموذج من نماذج معالجة التخمين وباختلاف عدد الفقرات وحجم عينة البيانات المولدة.

طول الاختبار	النموذج	الخطأ المعياري للتقدير ٥٠٠ مفحوص	الخطأ المعياري للتقدير ١٥٠٠ مفحوص
٢٥ فقرة	الثلاثي	٠,٠٦٠	٠,٠٢٦
	الثنائي المعدل	٠,٠٢٢	٠,٠٢٤
	عتبة التخمين	٠,١٣٧	٠,١٣٠
	القائم على الصعوبة	٠,١٨٨	٠,١٢٥
	الكفاءة النسبية لـ الثلاثي على الثنائي المعدل	٠,١٣	٠,٨٥
	الكفاءة النسبية لـ الثلاثي على عتبة التخمين	٥,٢١	٢٥
	الكفاءة النسبية لـ الثلاثي على القائم على الصعوبة	٩,٨٢	٢٣,١١
	الكفاءة النسبية لـ الثنائي المعدل على عتبة التخمين	٣٨,٧٨	٢٩,٣٤
	الكفاءة النسبية لـ الثنائي المعدل على القائم على الصعوبة	٧٣,٠٢	٢٧,١٣
	الكفاءة النسبية لـ عتبة التخمين على القائم على الصعوبة	١,٨٨	٠,٩٢
٥٠ فقرة	الثلاثي	٠,٠٢٢	٠,٠١٦
	الثنائي المعدل	٠,٠٢٠	٠,٠٢٠
	عتبة التخمين	٠,١٣٢	٠,٠٩٤
	القائم على الصعوبة	٠,١٣٨	٠,١١٥
	الكفاءة النسبية لـ الثلاثي على الثنائي المعدل	٠,٨٣	١,٥٦
	الكفاءة النسبية لـ الثلاثي على عتبة التخمين	٣٦,٠٠	٣٤,٥٢
	الكفاءة النسبية لـ الثلاثي على القائم على الصعوبة	٣٩,٠٠	٥١,٦٦
	الكفاءة النسبية لـ الثنائي المعدل على عتبة التخمين	٤٣,٥٦	٢٢,٠٩
	الكفاءة النسبية لـ الثنائي المعدل على القائم على الصعوبة	٤٧,٦١	٣٣,٠٦
	الكفاءة النسبية لـ عتبة التخمين على القائم على الصعوبة	١,٠٩	١,٥٠

يُلاحظ من الجدول (٣٧) ان النتائج الخاصة به قد جاءت على النحو الآتي:

اولاً: في حالة الاختبار المُكون (٢٥) فقرة والمُولد عليه (٥٠٠)استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة الصعوبة للنموذج اللوجستي الثلاثي المعلمة على نموذجي عتبة التخمين والقائم على الصعوبة قد كانت ظاهريًا لصالح النموذج اللوجستي الثلاثي المعلمة، بينما

كانت لصالح النموذج الثنائي المعدل عند المقارنة معه، كما كانت قيم الكفاءة النسبية لمعلمة الصعوبة لصالح النموذج الثنائي المعدل عند مقارنته مع كل من النموذج القائم على الصعوبة، ونموذج عتبة التخمين، وكانت قيم الكفاءة النسبية لمعلمة الصعوبة لصالح نموذج عتبة التخمين عند المقارنة مع النموذج القائم على الصعوبة.

ثانياً: في حالة الاختبار المُكون من (٥٠) فقرة والمُولد عليه (٥٠٠) استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة الصعوبة للنموذج اللوجستي الثلاثي المعلمة على نموذجي عتبة التخمين والقائم على الصعوبة قد كانت ظاهريًا لصالح النموذج اللوجستي الثلاثي المعلمة، بينما كانت لصالح النموذج الثنائي المعدل عند المقارنة معه، كما كانت قيم الكفاءة النسبية لمعلمة الصعوبة لصالح النموذج الثنائي المعدل عند مقارنته مع كل من النموذج القائم على الصعوبة، ونموذج عتبة التخمين، وكانت قيم الكفاءة النسبية لمعلمة الصعوبة لصالح نموذج عتبة التخمين عند المقارنة مع النموذج القائم على الصعوبة.

ثالثاً: في حالة الاختبار المُكون من (٢٥) فقرة والمولد عليه (١٥٠٠) استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة الصعوبة للنموذج اللوجستي الثلاثي المعلمة على نموذجي عتبة التخمين والقائم على الصعوبة قد كانت ظاهريًا لصالح النموذج اللوجستي الثلاثي المعلمة، بينما كانت لصالح النموذج الثنائي المعدل عند المقارنة معه، كما كانت قيم الكفاءة النسبية لمعلمة الصعوبة لصالح النموذج الثنائي المعدل عند مقارنته مع كل من النموذج القائم على الصعوبة ونموذج عتبة التخمين، وكانت قيم الكفاءة النسبية لمعلمة الصعوبة لصالح النموذج القائم على الصعوبة عند المقارنة مع نموذج عتبة التخمين.

رابعًا: في حالة الاختبار المُكون من (٥٠) فقرة والمولد عليه (١٥٠٠) استجابة لمفحوص:

ان قيم الكفاءة النسبية لمعلمة الصعوبة كانت لصالح النموذج اللوجستي الثلاثي المعلمة عند مقارنته مع النماذج الأخرى المستخدمة لمعالجة التخمين، وعند مقارنة قيم الكفاءة النسبية لمعلمة الصعوبة للنموذج الثنائي المُعدل مع كل من نموذج عتبة التخمين والنموذج القائم على الصعوبة كانت قيم الكفاءة النسبية لصالح النموذج الثنائي المعدل، وعند مقارنة نموذج عتبة التخمين مع النموذج القائم على الصعوبة كانت قيمة الكفاءة النسبية لصالح نموذج عتبة التخمين.

ثالثاً: إيجاد الكفاءة النسبية لمعلمة التخمين.

جدول (٣٨)

قيم الكفاءة النسبية لمعلمة التخمين وفقاً لكل نموذج من نماذج معالجة التخمين وباختلاف عدد الفقرات وحجم عينة البيانات المولدة.

طول الاختبار	النموذج	الخطأ المعياري للتقدير	
		٥٠٠ مفحوص	١٥٠٠ مفحوص
٢٥ فقرة	الثلاثي	٠,٠٧٦	٠,٠٥٨
	الثنائي المعدل	*	*
	عتبة التخمين	٠,٠٨٣	٠,٠٨٣
	القائم على الصعوبة	٠,١٠٩	٠,٠٧٥
	الكفاءة النسبية لـ الثلاثي على الثنائي المعدل	*	*
	الكفاءة النسبية لـ الثلاثي على عتبة التخمين	١,١٩	٢,٠٥
	الكفاءة النسبية لـ الثلاثي على القائم على الصعوبة	٢,٠٦	١,٦٧
	الكفاءة النسبية لـ الثنائي المعدل على عتبة التخمين	*	*
	الكفاءة النسبية لـ الثنائي المعدل على القائم على الصعوبة	*	*
	الكفاءة النسبية لـ عتبة التخمين على القائم على الصعوبة	١,٧٢	٠,٨٢
٥٠ فقرة	الثلاثي	٠,٥٤	٠,٠٣٠
	الثنائي المعدل	*	*
	عتبة التخمين	٠,٠٧٨	٠,٠٧٩
	القائم على الصعوبة	٠,٠٧٧	٠,٠٦٨
	الكفاءة النسبية لـ الثلاثي على الثنائي المعدل	*	*
	الكفاءة النسبية لـ الثلاثي على عتبة التخمين	٢,٠٩	٦,٩٣
	الكفاءة النسبية لـ الثلاثي على القائم على الصعوبة	٢,٠٣	٥,١٤
	الكفاءة النسبية لـ الثنائي المعدل على عتبة التخمين	*	*
	الكفاءة النسبية لـ الثنائي المعدل على القائم على الصعوبة	*	*
	الكفاءة النسبية لـ عتبة التخمين على القائم على الصعوبة	٠,٩٧	٠,٧٤

* تم تثبيت معلمة التخمين عند القيمة ٠,٢٠

يُلاحظ من الجدول (٣٨) ان النتائج الخاصة به قد جاءت على النحو الآتي :

اولاً: في حالة الاختبار المُكون من (٢٥) فقرة والمولد عليه (٥٠٠) استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة التخمين قد كانت ظاهريًا لصالح النموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النماذج الأخرى المستخدمة لمعالجة التخمين، واخيرًا كانت قيم الكفاءة النسبية لمعلمة التخمين لصالح نموذج عتبة التخمين عند المقارنة مع النموذج القائم على الصعوبة.

ثانياً: في حالة الاختبار المُكون من (٥٠) فقرة والمولد عليه (٥٠٠) استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة التخمين قد كانت ظاهريًا لصالح النموذج اللوجستي الثلاثي المعلمة عند المقارنة مع كل من نموذج عتبة التخمين والنموذج القائم على الصعوبة، واخيرًا كانت قيم الكفاءة النسبية لمعلمة التخمين لصالح النموذج القائم على الصعوبة عند المقارنة مع نموذج عتبة التخمين.

ثالثاً: في حالة الاختبار المُكون من (٢٥) فقرة والمولد عليه (١٥٠٠) استجابة لمفحوص:

يُلاحظ أن قيم الكفاءة النسبية لمعلمة التخمين قد كانت لصالح النموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النماذج الأخرى المستخدمة لمعالجة التخمين، واخيرًا كانت قيم الكفاءة النسبية لمعلمة التخمين لصالح النموذج القائم على الصعوبة عند المقارنة مع نموذج عتبة التخمين.

رابعًا: في حالة الاختبار المُكون من (٥٠) فقرة والمولد عليه (١٥٠٠) استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة التخمين قد جاءت لصالح النموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النماذج الأخرى المستخدمة لمعالجة التخمين، كما كانت قيم الكفاءة النسبية لمعلمة التخمين لصالح النموذج القائم على الصعوبة عند المقارنة مع نموذج عتبة التخمين.

رابعًا: إيجاد الكفاءة النسبية لمعلمة القدرة للمفحوصين.

جدول (٣٩)

قيم الكفاءة النسبية لمعلمة القدرة وفقاً لكل نموذج من نماذج معالجة التخمين وباختلاف عدد الفقرات وحجم عينة البيانات المولدة.

طول الاختبار	النموذج	الخطأ المعياري للتقدير ٥٠٠ مفحوص	الخطأ المعياري للتقدير ١٥٠٠ مفحوص
٢٥ فقرة	الثلاثي	٠,٢٨٥	٠,٢٣٧
	الثنائي المعدل	٠,٢٩٣	٠,٢٤٤
	عتبة التخمين	٠,٣٩٥	٠,٢٤١
	القائم على الصعوبة	٠,٣٩٦	٠,٣٠٢
	الكفاءة النسبية لـ الثلاثي على الثنائي المعدل	١,٠٦	١,٠٦
	الكفاءة النسبية لـ الثلاثي على عتبة التخمين	١,٩٢	١,٠٣
	الكفاءة النسبية لـ الثلاثي على القائم على الصعوبة	١,٩٣	١,٦٢
	الكفاءة النسبية لـ الثنائي المعدل على عتبة التخمين	١,٨٢	٠,٩٨
	الكفاءة النسبية لـ الثنائي المعدل على القائم على الصعوبة	١,٨٣	١,٥٣
	كفاءة النسبية لـ عتبة التخمين على القائم على الصعوبة	١,٠١	١,٥٧
٥٠ فقرة	الثلاثي	٠,١٣٢	٠,١٥٣
	الثنائي المعدل	٠,١٣٢	٠,١٥٩
	عتبة التخمين	٠,١٣١	٠,١٦٠
	القائم على الصعوبة	٠,٢٨٥	٠,١٨٦
	الكفاءة النسبية لـ الثلاثي على الثنائي المعدل	١	١,٠٨
	الكفاءة النسبية لـ الثلاثي على عتبة التخمين	٠,٩٨	١,٠٩
	الكفاءة النسبية لـ الثلاثي على القائم على الصعوبة	٤,٦٦	١,٤٨
	الكفاءة النسبية لـ الثنائي المعدل على عتبة التخمين	٠,٩٨	١,٠١
	الكفاءة النسبية لـ الثنائي المعدل على القائم على الصعوبة	٤,٦٦	١,٣٧
	الكفاءة النسبية لـ عتبة التخمين على القائم على الصعوبة	٤,٧٣	١,٣٥

يُلاحظ من الجدول (٣٩) أن النتائج الخاصة به قد جاءت على النحو الآتي :

اولاً: في حالة الاختبار المكون من (٢٥) فقرة والمولد عليه (٥٠٠) استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة القدرة للمفحوصين قد جاءت لصالح النموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النماذج الأخرى، وكانت قيم الكفاءة النسبية لصالح النموذج الثنائي المعدل عند المقارنة مع النموذج القائم على الصعوبة ونموذج عتبة التخمين، وعند مقارنة نموذج عتبة التخمين مع النموذج القائم على الصعوبة كانت قيم الكفاءة النسبية لصالح نموذج عتبة التخمين.

ثانيًا: في حالة الاختبار المكون من (٥٠) فقرة والمولد عليه (٥٠٠) استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة القدرة للمفحوصين قد كانت لصالح النموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النموذج القائم الصعوبة في حين كانت لصالح نموذج عتبة التخمين عند المقارنة معه، وعند المقارنة مع النموذج الثنائي المعدل كانت قيمة الكفاءة النسبية مساوية للواحد، وكانت قيم الكفاءة النسبية لصالح نموذج عتبة التخمين عند مقارنته مع النموذج الثنائي المعدل، في حين كانت قيم الكفاءة النسبية لصالح النموذج الثنائي المعدل عند المقارنة مع النموذج القائم على الصعوبة، واخيرًا كانت قيم الكفاءة النسبية لمعلمة القدرة للمفحوصين لصالح نموذج عتبة التخمين عند المقارنة مع النموذج القائم على الصعوبة.

ثالثًا: في حالة الاختبار المكون من (٢٥) فقرة والمولد عليه (١٥٠٠) استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة القدرة للمفحوصين قد كانت لصالح النموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النماذج الأخرى المستخدمة لمعالجة التخمين، وكانت قيم الكفاءة النسبية لصالح نموذج عتبة التخمين عند مقارنته مع النموذج الثنائي المعدل، في حين كانت قيم الكفاءة النسبية لصالح النموذج الثنائي المُعدل عند المقارنة مع النموذج القائم على الصعوبة، واخيرًا كانت قيم الكفاءة النسبية لمعلمة القدرة للمفحوصين لصالح نموذج عتبة التخمين عند المقارنة مع النموذج القائم على الصعوبة.

رابعًا: في حالة الاختبار المكون من (٥٠) فقرة والمولد عليه (١٥٠٠) استجابة لمفحوص:

أن قيم الكفاءة النسبية لمعلمة القدرة للمفحوصين قد جاءت لصالح النموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النماذج الأخرى، وكانت قيم الكفاءة النسبية لصالح النموذج الثنائي المعدل عند المقارنة مع النموذج القائم على الصعوبة ونموذج عتبة التخمين، وعند مقارنة نموذج عتبة التخمين مع النموذج القائم على الصعوبة كانت قيم الكفاءة النسبية لصالح نموذج عتبة التخمين.

الفصل الخامس

مناقشة النتائج والتوصيات

هدفت هذه الدراسة إلى التحقق من فاعلية اربعة نماذج من نظرية الاستجابة للفقرة ضمن إطار النموذج اللوجستي الثلاثي المعلمة في معالجة التخمين في الاختبارات الموضوعية/الاختيار من متعدد وتحري أثر هذه النماذج على دقة التقدير لكل من معالم الفقرة ومعلمة القدرة للمفحوصين باختلاف عدد فقرات الاختبار (٢٥ فقرة و٥٠ فقرة) وحجم عينة المفحوصين (٥٠٠ مفحوص و١٥٠٠ مفحوص).

ويتناول هذا الفصل عرضًا لمناقشة النتائج والتوصيات التي توصلت لها هذه الدراسة مرتبة على النحو الآتي:

السؤال الأول: هل يختلف تقدير معلمة القدرة بإختلاف النموذج المستخدم لمعالجة التخمين في إطار النموذج اللوجستي الثلاثي المعلمة تبعًا لمتغيري طول الاختبار (٢٥ فقرة و٥٠ فقرة) وعدد المفحوصين (٥٠٠ مفحوص و١٥٠٠ مفحوص).

أظهرت النتائج التي انتهت إليها الدراسة الحالية فيما يتعلق بتقدير معلمة القدرة باختلاف النموذج المستخدم لمعالجة التخمين وعدد فقرات الاختبار وحجم عينة المفحوصين أن الفروق في المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ ولمؤشر التحيز كانت دالة إحصائيا لصالح النموذج اللوجستي الثلاثي المعلمة، يليه النموذج الثنائي المعدل، ثم نموذج عتبة التخمين، واخيرًا النموذج القائم على الصعوبة.

كما أظهرت النتائج أن متوسطات الجذر التربيعي لمتوسط مربع الخطأ تقل بزيادة طول الاختبار من (٢٥) فقرة إلى (٥٠) فقرة، وكذلك أن الدقة في تقدير معلمة القدرة تزداد بزيادة عدد المفحوصين من (٥٠٠) مفحوص إلى (١٥٠٠) مفحوص، واتفقت هذه النتائج مع نتائج دراسة فو (Fu, 2010) والتي أشارت إلى أن النموذج اللوجستي الثلاثي المعلمة ونموذج تثبيت معلمة التخمين، يُعطيان تقديرات دقيقة لمعالم الفقرة ومعلمة القدرة حينما يكون هناك مواءمة جيدة بين كل من معلمة الصعوبة للفقرة ومعلمة القدرة للمفحوص، كما اتفقت نتائج الدراسة الحالية مع دراسة عبابنه (٢٠٠٤) والتي أشارت إلى أن دقة التقديرات لمعالم الفقرات ومعلمة القدرة تزداد بزيادة حجم عينة المفحوصين وبزيادة طول الاختبار، كما اتفقت نتائج هذه الدراسة نسبيا مع دراسة الحاج (٢٠١١)

والتي أشارت إلى أن دقة التقديرات لكل من معالم الفقرة ومعلمة القدرة للمفحوصين أفضل في حالة النموذج اللوجستي الثلاثي المعلمة منها في حالة النموذج الثنائي المعدل، كما اتفقت نتائج الإجابة عن هذا السؤال مع دراسة ستيدي (Setiadi ,1997) والتي أشارت إلى ان دقة التقدير لمعالم الفقرات كانت افضل في حالة استخدام النموذج اللوجستي الثلاثي المعلمة مقارنة مع النماذج الأخرى التي استخدمها الباحث وهي النموذج الأحادي المعلمة والنموذج الأحادي المعلمة المعدل، كما اتفقت النتائج التي توصلت لها هذه الدراسة مع ما أشارا اليه هامبلتون وكوك(Hambleton and Cook, 1983) من أن حجم عينة المفحوصين المستخدمة في تقدير معالم الفقرة ومعلمة القدرة يؤثر على استقرار ودقة معالم الفقرات ومعلمة القدرة للمفحوصين، كما أظهرت نتائج تحليل مؤشر التحيز (Bias) تشابهًا مع نتائج الجذر التربيعي لمتوسط مربع الخطأ (RMSE) في تقدير معلمة القدرة، وفي سياق الإجابة عن هذا السؤال تم إيجاد المتوسطات الحسابية للخطأ المعياري للتقدير لمعلمة القدرة باختلاف عدد فقرات الاختبار وحجم عينة المفحوصين.

وقد أظهرت نتائج هذه المتوسطات الحسابية انخفاضًا ظاهريًا في قيم المتوسطات الحسابية للخطأ المعياري للتقدير بزيادة حجم عينة المفحوصين من (٥٠٠) مفحوص إلى (١٥٠٠) مفحوص وبزيادة عدد فقرات الاختبار من (٢٥) فقرة إلى (٥٠) فقرة، وهذا يتفق مع ما أشار إليه ديمارس (Demars,2010) من أن زيادة عدد فقرات الاختبار وزيادة حجم عينة المفحوصين تقلل من قيم الخطأ المعياري للتقدير، كما تقل قيم الخطأ المعياري للتقدير باقتراب معلمة الصعوبة للفقرة من معلمة القدرة للمفحوصين أي حينما تكون هناك مواءمة جيدة بينهما، وقد أظهرت النتائج أن المتوسطات الحسابية للخطأ المعياري للتقدير لمعلمة القدرة كان اقل في حالة استخدام النموذج اللوجستي الثلاثي المعلمة عندما يكون عدد فقرات الاختبار (٥٠) فقرة وحجم عينة المفحوصين (١٥٠٠) مفحوص.

كما اتفقت نتائج هذه الدراسة نسبيًا مع دراسة جاو (Gao, 2011) والتي أشارت الى أن دقة تقديرات معلمة الصعوبة في حالة النموذج اللوجستي الثلاثي المعلمة أفضل منها في حالة المقارنة مع النماذج الأخرى المشمولة في الدراسة، كما اتفقت نتائج هذه الدراسة نسبيًا مع دراسة ستيزر (Stetzer, 2007) والتي أظهرت أن دقة التقدير لمعالم الفقرات ومعلمة القدرة في حالة استخدام النموذج اللوجستي الثلاثي المعلمة افضل منها عند استخدام نموذج توسط الجهد الذي استخدمته تلك الدراسة والذي تم الإشارة اليه خلال استعراض الادب النظري في الفصل الثاني من هذه الدراسة.

كما اتفقت نتائج الدراسة الحالية مع نتائج دراسة والر (Waller, 1989) والتي أشارت إلى أن دقة التقديرات لمعالم الفقرة ومعلمة القدرة للنموذج اللوجستي الثلاثي المعلمة أفضل منها في حالة استخدام نموذج (إزالة التخمين العشوائي) والنموذج اللوجستي الثلاثي المعلمة ذي معلمة التخمين الثابتة.

واختلفت نتائج هذه الدراسة مع نتائج دراسة سوبادي (Subedi, 2009) والتي أظهرت أن دقة التقديرات لكل من معالم الفقرة ومعلمة القدرة كانت أفضل عند استخدام نموذجين مختلطين مقارنة بكل من النموذج اللوجستي الثلاثي المعلمة والنموذج اللوجستي الثنائي المعلمة، كما اختلفت نتائج هذه الدراسة مع نتائج دراسة (Pelton, 2002) والتي توصلت إلى أن النموذج اللوجستي الثنائي المعلمة المُعدّل يُقدم تقديرات لمعلمة الصعوبة للفقرة أكثر دقة من تلك التي يُقدمها نموذج راش والنموذج اللوجستي الثلاثي المعلمة .

ويمكن أن يُعزى تفوق النموذج اللوجستي الثلاثي المعلمة على النموذج الثنائي المُعدّل بسبب إضافة ثابت التخمين (0.20) للنموذج الثنائي المُعدّل؛ الأمر الذي يُعّد كأنه خطأ ثابت تمت أضافته للنموذج وهذا يقود إلى أن تكون دقة التقديرات في النموذج الثنائي المُعدّل أقل من دقة التقديرات في النموذج اللوجستي الثلاثي المعلمة؛ وهذا بالتالي يُفسر أن أخطاء التقدير لمعلمة الفقرة او للمفحوصين أعلى في حالة النموذج الثنائي المُعدّل منها في حالة النموذج اللوجستي الثلاثي المعلمة، ممّا يُفسّر أيضًا ان تكون الكفاءة النسبية لمعالم الفقرة أو لمعلمة القدرة أيضًا لصالح النموذج اللوجستي الثلاثي المعلمة مقارنة مع النماذج الأخرى المستخدمة في إطار هذه الدراسة لمعالجة التخمين.

كما يمكن تفسير دقة التقديرات لمعلمة القدرة أو لمعالم الفقرات في حالة النموذج اللوجستي الثلاثي المعلمة مقارنة بنموذج عتبة التخمين والنموذج القائم على الصعوبة؛ بسبب وجود معالم إضافية في هذه النماذج وبالتالي حاجتها إلى زمن أطول لإجراء عملية التقدير مقارنة مع النماذج الأخرى كما أشار إلى ذلك كل من فرايك وستروبل وليسش وزيليس (Frick, Strobl, Leish, and Zeileis, 2012)؛ كما أن هذه النماذج تتأثر بشكل كبير بالمعلومات القبلية (Priors Information) التي يتم تزويد البرنامج الحاسوبي المسُتخدم بها لإتمام عملية التقدير والتي تؤثر بشكل كبير على دقة التقديرات لكل من معالم الفقرة ومعلمة القدرة للمفحوصين لهذه النماذج.

السؤال الثاني:

هل يختلف تقدير معالم الفقرات باختلاف النموذج المســـتخدم لمعالجة التخمين في إطار النموذج الثلاثي المعلمة تبعًا لمتغيري طول الاختبار (٢٥ فقرة و٥٠ فقرة) وعدد المفحوصـــين (٥٠٠ مفحوص و١٥٠٠ مفحوص).

اولاً معلمة التمييز:

أظهرت النتائج افضلية للنموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النماذج الاخرى المستخدمة لمعالجة التخمين، واظهر النموذج الثنائي المعدل أفضلية على النموذج القائم على الصعوبة ولم تكن الفروق دالة مع نموذج عتبة التخمين، واظهر نموذج عتبة التخمين أفضلية على النموذج القائم على الصعوبة، كما اظهرت النتائج انخفاضًا في المتوسطات الحسابية لقيم الجذر التربيعي لمتوسط مربع الخطأ ظاهريًا بزيادة عدد فقرات الاختبار من (٢٥) فقرة إلى (٥٠) وبزيادة حجم عينة المفحوصين من (٥٠٠) مفحوص إلى (١٥٠٠) مفحوص على التوالي، إلا أن هذه الفروق لم تكن ذات دلالة احصائية عند مستوى الدلالة (α= 0.05).

ثانياً: معلمة الصعوبة:

اظهرت نتائج الدراسة وجود فروق دالة إحصائياً عند مستوى الدلالة (α=(0.05 بين المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ تُعزى للنموذج المستخدم لمعالجة التخمين وذلك لصالح النموذج اللوجستي الثلاثي المعلمة، ثم للنموذج الثنائي المعدل على كل من نموذج عتبة التخمين والنموذج القائم على الصعوبة، ولم تكن الفروق ذات دلالة عند مقارنة نموذج عتبة التخمين مع النموذج القائم على الصعوبة، كما اظهرت المتوسطات الحسابية لقيم الجذر التربيعي لمتوسط مربع الخطأ انخفاضًا ظاهريًا في قيمها لجميع النماذج المستخدمة لمعالجة التخمين بزيادة حجم عينة المفحوصين من (٥٠٠) مفحوص إلى (١٥٠٠) مفحوص وبزيادة عدد فقرات الاختبار من (٢٥) فقرة إلى (٥٠) فقرة.

ثالثاً: معلمة التخمين:

اظهرت النتائج التي توصلت لها هذه الدراسة فيما يتعلق بدقة تقدير معلمة التخمين وجود فروق دالة إحصائيًا بين المتوسطات الحسابية للجذر التربيعي لمتوسط مربع الخطأ تُعزى للنموذج المستخدم لمعالجة التخمين وذلك لصالح النموذج اللوجستي الثلاثي المعلمة،ثم للنموذج القائم على الصعوبة ثم لنموذج عتبة التخمين كما اظهرت المتوسطات الحسابية لقيم الجذر التربيعي لمتوسط مربع الخطأ

انخفاضًا ظاهريًا في قيمها بزيادة حجم عينة المفحوصين من (٥٠٠) مفحوص على (١٥٠٠) مفحوص وبزيادة عدد فقرات الاختبار من (٢٥) فقرة إلى (٥٠) فقرة.

اتفقت النتائج التي توصلت لها الدراسة في الإجابة عن هذا السؤال مع نتائج دراسة العبابنه(٢٠٠٤) ومع نتائج دراسة الحاج(٢٠١١) والتي أشارت إلى أن دقة التقدير لمعالم الفقرة ومعلمة القدرة للمفحوصين كانت افضل في حالة النموذج اللوجستي الثلاثي المعلمة مقارنة مع النموذج الثنائي المعدل، كما اتفقت نتائج هذه الدراسة مع نتائج دراسة بارشل وكرومي وجيسون (Parshall, Krromrey, Chason and Yi, 1997) ، كما اتفقت نتائج الدراسة مع ما أشار اليه كل من سوامنثان جيفورد (Swaminathan and Gifford, 1983) و هامبلتون وكوك (Hambleton and Cook, 1983) و وهامبلتون وجونز(Hambelton and Jones, 1994) والتي أشارت جميعها إلى زيادة الدقة في تقدير معالم الفقرات بزيادة حجم عينة المفحوصين وزيادة عدد فقرات الاختبار.

السؤال الثالث:

هل تختلف قيم الكفاءة النســبية لمعالم الفقرات ومعلمة القدرة بإختلاف متغيرات الدراســة (النموذج المستخدم لمعالجة التخمين وعدد الفقرات وحجم العينة)؟

اولاً: معلمة التمييز.

أظهرت نتائج مؤشر الكفاءة النسبية في تقدير معلمة التمييز افضلية للنموذج اللوجستي الثلاثي المعلمة عند المقارنة مع النماذج الأخرى المستخدمة لمعالجة التخمين يليه النموذج الثنائي المعدل، ثم نموذج عتبة التخمين، واخيرًا النموذج القائم على الصعوبة.

ثانياً: معلمة الصعوبة.

اظهرت النتائج أن قيم الكفاءة النسبية في تقدير معلمة الصعوبة هي لصالح النموذج اللوجستي الثلاثي المعلمة يليه النموذج الثنائي المعدل ثم النموذج القائم على الصعوبة، واخيرًا نموذج عتبة التخمين.

ثالثاً: معلمة التخمين:

أظهرت النتائج المتعلقة بالإجابة عن هذا السؤال أن قيم الكفاءة النسبية قد كانت لصالح النموذج اللوجستي الثلاثي المعلمة عند مقارنته مع كل من نموذج عتبة التخمين والنموذج القائم على الصعوبة، وجاءت قيم الكفاءة النسبية لصالح نموذج عتبة التخمين في حالة الاختبار المكون من

(٢٥) فقرة والعينة المكونة من (٥٠٠) مفحوص، وفي حالة الاختبار المكون من (٥٠) فقرة كانت قيم الكفاءة النسبية لصالح النموذج القائم على الصعوبة.

ويمكن تفسير نتائج الإجابة عن سؤال الدراسة الثالث من خلال العلاقة التي تربط بين المؤشرات: الجذر التربيعي لمتوسط مربع الخطأ والتحيز والخطأ المعياري للتقدير $RMSE^2=Bias^2+SEE^2$ وبالتالي فإن القيمة الكبيرة للجذر التربيعي لمتوسط مربع الخطأ تترافق مع قيمة كبيرة للخطأ المعياري للتقدير والعكس صحيح وبالتالي فإن النماذج التي كانت متوسطات الجذر التربيعي لمتوسط مربع الخطأ لها كبيرة فإن الخطأ المعياري للتقدير لها سيكون مرتفعًا أيضًا (Wang and Wang,2002; Weiss and Babcock, 2009).

رابعًا: معلمة القدرة.

اظهرت نتائج الإجابة عن هذا السؤال ان قيم الكفاءة النسبية كانت لصالح النموذج اللوجستي الثلاثي المعلمة عند مقارنته مع النماذج الاخرى المستخدمة لمعالجة التخمين في هذه الدراسة يليه النموذج الثنائي المعدل ثم نموذج عتبة التخمين ثم النموذج القائم على الصعوبة.

وبشكل عام يمكن القول إن عملية مقارنة نتائج اسئلة هذه الدراسة مع نتائج الدراسات السابقة ذات الصلة والواردة في الفصل الثاني من هذه الدراسة، فيها شيء من الصعوبة وذلك بسبب اختلاف النماذج المستخدمة واختلاف متغيرات كل دراسة، كما أن البرمجيات الحاسوبية التي اسُتخدمت لتقدير معالم الفقرات ومعلمة القدرة للمفحوصين تختلف من دراسة لأخرى، وفي حالة الدراسات التي استخدمت (المحاكاة) اسلوب توليد البيانات فمجرد الاختلاف في خصائص التوزيعات لمعلمة القدرة ومعالم الفقرات (صعوبة وتمييز وتخمين) للبيانات المولدة يجعل عملية المقارنة فيها شيء من الصعوبة، عدا عن أن كثيرًا من الدراسة المشابهة اعتمدت على الاحصاء الوصفي فقط في الاجابة عن أسئلتها، بينما استندت هذه الدراسة بشكل أساسي للاحصاء التحليلي.

التوصيات:

بعد استعراض ومناقشة النتائج التي توصلت إليها هذه الدراسة والإجابة عن أسئلتها، وكغالبية الدراسات التي تسعى للمقارنة بين فعالية النماذج المختلفة في التغلب على مشكلة التخمين العشوائي في إطار نظرية الاستجابة للفقرة وأثرها على دقة التقديرات لمعالم الفقرات ومعلمة القدرة للمفحوصين؛ فإن هناك حاجة إلى المزيد من الدراسات التي تراعي ظروفًا ومتغيرات مختلفة لتأكيد افضلية نموذج ما على نموذج آخر.

وفي ضوء ذلك فإن الباحث يوصي بما يلي:

- يمكن للباحثين مواصلة البحث في مقارنة هذه النماذج بظروف أخرى مختلفة من خلال استخدام تشكيلات مختلفة من حجم العينة وطول الاختبار غير تلك الواردة في الدراسة.
- اجراء دراسات تضم بيانات تجريبية مستمدة من الواقع التجريبي وبيانات يتم توليدها باستخدام أسلوب المحاكاة.
- استخدام النموذج اللوجستي الثلاثي المعلمة؛ حيث انه اظهر أفضلية على النماذج الأخرى المستخدمة في هذه الدراسة.
- دراسة أثر توليد معالم الفقرات ومعلمة القدرات بتوزيعات أخرى مختلفة، ففي هذه الدراسة تم توليد معلمة القدرة بتوزيع طبيعي بمتوسط (صفر) وانحراف معياري (واحد) وكذلك معلمة الصعوبة ويمكن استخدام التوزيع المنتظم لكل من القدرة ومعلمة الصعوبة وتحري أثره على دقة التقدير سواء لمعالم الفقرات أو الأفراد.
- المقارنة بين فعالية هذه النماذج لمعالجة مشكلة التخمين من خلال استخدام طريقتي التقدير البييزية والارجحية العظمى بنفس الدراسة.
- استخدام أكثر من قيمة كثابت للتخمين للنماذج المعدلة حيث أن هذه الدراسة استخدمت قيمة ثابتة تساوي (٠,٢٠).
- استخدام طريقة التقدير البييزية في تدريج فقرات بنوك الأسئلة لما لها من دقة في تقدير معلمة القدرة ومعالم الفقرات بسبب وجود توزيعًات سابقة لمعالم الفقرات وقدرات المفحوصين يتم تزويدها للبرامج الحاسوبية المستخدمة.
- المقارنة بين إجراءات معالجة التخمين في إطار النظرية الكلاسيكية للقياس ونظرية الاستجابة للفقرة.
- المقارنة بين نتائج النماذج المستخدمة في هذه الدراسة لمعالجة التخمين باستخدام أكثر من برنامج حاسوبي.
- ادخال متغير الجنس (ذكر وانثى) والبيئة في دارسة أثر التخمين على دقة تقدير معالم الفقرات ومعلمة القدرة للمفحوصين.
- إجراء دراسات مماثلة واستخدام نماذج معدلة بحيث يكون التعديل فيها على معلمة تمييز الفقرة.

المراجع العربية

البطش، محمد وليد.(٢٠٠٥). الاتجاهات الحديثة في مجال القياس والتقويم وتطبيقاتها في ميدان التربية الخاصة. **ورقة مقدمة إلى مؤتمر التربية الخاصة العربي : الواقع والمأمول.** الجامعة الأردنية، عمان، الأردن، ٢٦- ٢٠٠٥/٤/٢٧.

العبابنة، عماد غصاب.(٢٠٠٤). **أثر حجم العينة وطريقة إنتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة.** أطروحة دكتوراه غير منشورة، جامعة عمان للدراسات العليا،عمّان ،الاردن.

الخرشا، طه .(٢٠٠٤). **مقارنة ثلاثة طرق للتصحيح ضد أثر التخمين في النظرية الكلاسيكية.** أطروحة ماجستير غير منشورة، جامعة مؤته،الكرك، الاردن.

الزبون، حابس.(٢٠٠٤). **أثر استخدام ثلاث طرق للتصحيح لضبط أثر التخمين في الخصائص السيكومترية لاختبارات الاختيار من متعدد في ضوء نظرية الاستجابة للفقرة.** أطروحة ماجستير غير منشورة.جامعة مؤته، الكرك، الأردن.

الحاج، احمد بكر.(٢٠١١). **أثر التحكم بمعلمة التخمين في نماذج استجابة الفقرة على تقدير معالم الفقرات والأفراد.** أطروحة دكتوراه غير منشورة، جامعة اليرموك، اربد، الأردن.

راشد، انور محمد.(٢٠١١). **أثر تعليمات التخمين في استجابة الطلبة لفقرات الاختيار من متعدد على درجاتهم وعلى الخصائص السيكومترية للاختبار وفقراته.** أطروحة ماجستير غير منشورة، جامعة عمان العربية.عمّان،الاردن.

الشواورة، شادي.(٢٠١٣).**دقة تقدير معالم الفقرات بطريقتي الارجحية العظمى الهامشية وبييز في ظروف مختلفة في عدد الفقرات وحجم العينة والنموذج اللوغاريتمي المستخدم.** أطروحة دكتوراه غير منشورة. جامعة اليرموك، اربد.الاردن.

الشريفين، نضال.(٢٠١٢). أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. **المجلة التربوية،**١٢ (١٠٤)، ٢٢-٦٧.

عبد الوهاب، محمد محمود.(٢٠١٣). المقارنة بين النظرية الكلاسيكية والنموذج الثلاثي البارامتر في معالجة أثر التخمين عند تقدير إحصائيات الأفراد والمفردات في اختبار اختيار من متعدد. **المجلة التربوية**، الكويت، ٢٨(١٠٩)، ١٦٩-٢١٠.

المراجع الاجنبية

Alnabhan, Mousa .(2002).An Empirical Investigation of the Effect of three Methods of Handling Guessing and Risk Taking on the Psychometric Properties of a Test. **Social Behavior and Personality**, 30, 645-652.

Abu-Sayf, F.K.(1977). A New Formula Score. **Educational and Psychological Measurement**. 37, 853-862.

Adams, Ray. And Wu, Margaret.(2000). Pisa 2000 Technical Report. **Organisation for Economic Co-oporeation and Develpoment** ,OCED 2002.

Albanese, Mark A. and Forsyth Robert A.(1984). The One, Two and Modified Two-Parameter latent Trait Models: An Empirical Study of Relative Fit. **Educational And Psychological Measurement**, 44,229- 241.

Arnold, Taylor B. and Emerson, John.(2011). Nonparametric Goodness-of- Fit Tests for Discrete null Distributions.**The R Journal**. 3(2), 34-39.

Annie, W.Y and Chan, Alan H.S.(2012). Evaluation of the Three Multiple-Choice Assessment Methods in a human Factors Engineering Course. **Journal of the Chinese Institute of Industrial Engineers**, 29(7), 466-476.

Andrich ,David.(2012).Using Theorm by Andersen and the Dichotomous Rasch Model to assess the Presence of Random Guessing in multiple choice items. **Journal of Education Measurement**,37,417-442.

Baker ,FRANK B.(2004). **The Basic of Item Response Theory**. ERIC Clearinghouse on Assessment and Evaluation.

Bachman,L. Palmer,A.S.(1996).**Language testing in practice**. Oxford University Press.England.

Baldwin, Peter.(2006). A modified IRT Model Intended to Improve Parameter Estimates under Small Sample Conditions. **Paper Presented at the annual meeting of the National Council on Measurement in education**, San Franciso, CA: April 8-10.

Barnes, Laura L.B.and Wise Steven L.(1991). The Utility of a Modified One-Parameter IRT Model With Small Samples. **Applied Measurement In Education**, 4(2), 143-157.

Bar-Hillel, Maya. Budescu, David and Attali,Yigal.(2005). Scoring and keying Multiple Choice Tests: A case Study in Irrationality. **Mind & Society**, 3-12.

Bellhouse, D.R.(2004). The reverend Thomas Bayes FRS: A Biography to Celebrate The Tercentenary of his Birth. **Statistical Science**, 19(1), 3-43.

Ben, Shakhe.G.and Sinai ,Y.(1991).Gender Differences in Multiple-choice :the role of Differential guessing tendencies. **Journal of Educational Measurement**, 10,37-45.

Ben-Simon, A. Budescu ,D. and Nevo , B.(1997). A Comparative Study of Measures of Partial Knowledge in Multiple – Choice Tests. **Applied psychological Measurement**. 21, 65-88. ,

Blis, LB.(1980). A test of Lord's assumption regarding examinee Guessing behavior on multiple choice tests using elementary school students. **Journal of Education Measurement**, 17,53-147.

Boeck, De Paul. San Martin,Ernesto. and Del Pino Guido.(2006). IRT Models for Ability-Based Guessing**. Applied Psychological Measurement**, 30(3), 183-203.

Bovaird,J.A.(2002).**New Application in Testing: Using Response Time to Increase the Construct Validity of latent Trait Estimate**. Unpublished Doctoral Dissertation, University of Madson ,USA.

Bolger, Niall. And Kellagan, Thomas.(1990). Methods of Measurment and Gender Differences in Scholatic Achievement. **Journal of Education Measurement**, 27(2), 165-174.

Budescu D, and Bar-Hillel,M.(1993).To Guess or not to Guess: A Decision view of Formula Scoring Theoretic. **Journal of Educational Measurement**. 30,277-291.

Bush, m.(2001). A Multiple Choice Test that Rewards Partial Knowledge. **Journal of Further and Higher Education**, 25,23-44.

Burton, Richard F.(2002). Misinformation, Partial Knowledge and Guessing in true/ False Tests. **Medical education**, 36: 805-811.

Camilli, G.(2006).Test Fairness. In R. L. Bernnan(Ed), **Educational Measurment**. 4th .Westport, CT:Praeger.

Chiu,Twing-Wei.(2012).**Two new IRT Based Correction for Guessing.** Unpublished Doctoral Dissertation University of Colorado,USA.

Cao, J. and Stokes,S.L.(2008).Bayesian IRT Guessing Models for Partial Guessing Behavior. **Psychometrika**, 73,209-230.

Choppin, B.(1975). Guessing The Answer on Objective Tests. **Br.J.Educ Psychol**. 45, 206-213.

Coombs, C.H. Millhollands, J.E and Womer, F.B.(1956). **The assessment of partial knowledge. Educational and Psychological Measurement** . 16, 13-37.

Crocker,Linda and Algina, James.(1986).**Introduction to Classical and Modern Test Theory**. Holt, Rinehart and Winston,USA.

De Ayala, R.J.(2009).**The Theory And Practice Of Item Response Theory**. Guilford Press. New York, USA.

Demars, Christine.(2010).**Item Response Theory**. Oxford University Press, Inc.USA.

Dooden, Hamzeh M.(2005). Using Distractors in Correcting for Guessing in Multiple-Choice Tests. Dirasat, Educational Sgciences. 32(1), 192-198.

Dennick, Reg. and Tavakol, Mohsen.(2011). Making sense of Cronback's Alpha. **International Journal of Medical Education**, 2, 53-55.

Dawber, Teresa. Rogers, W.Todd and Carbonaro, Michael.(2009). Robustness of Lord's Formulas for Item Difficulty and Discrimination Conversions Between Classical and Item Theory Models.**The Alberta Journal of Educational research**, 4(55), 512-533.

Ebel, R.L. and David A.F.(1991) .**Essentials of Educational Measurement**. (3rd) ,USA, New Jersey: Prentice-Hall.

Espinosa, Maria Paz. and Gradezabal, Javier.(2010). Optimal Correction for Guessing in Multiple-Choice Tests. Retrieved Oct. 26.2013. from **http:// www.researchgate.net**

FU, Qiong.(2010). **Comparing Accuracy of parameter Estimation Using IRT Models in The Presence of Guessing**. Unpublished Doctoral Dissertation University of Illinois at Chicago,USA.

Fox, Jean-Paul.(2010).**Bayesian Item Response Modeling Theory and applications**. Springer, New York, USA.

Fray, Robert B.(1988). Formula Scoring of multiple-Choice tests(Correction for guessing). **Educational Measurement: Issues and Practice**. Summer 1988,

Frick, Hannah. Strobl, Carolin. Leish, Friedrich. and Zeileis, Achim.(2012). Flexible Rasch Mixture Models with Package Psychomix. **Journal of Statistical Software**, 48(7), 1-25.

Gao, Song.2011.**The Exploration of the Relationship Between Guessing and Latent Ability in IRT Models**. Unpublished Doctoral Dissertation University of southern Illinois ,USA.

Gregory ,Robert J.(1996).**Psychological Testing: History, Principles ,and Applications**. USA: Congress Catalogue-in-Publication Data.

Gruijtr, Dato N.M. and Kamp Leo J. Th.(2002). Statistical Test Theory For Education and Psychology. Retrieved May 10 2014 from: **http://irt.com.ne.kr**.

Kurz ,Terri Barber .(1999).A Review of Scoring Algorithms for Multiple-Choice Tests. **Paper Presented at the annual meeting of the Southwest educational Research Association**, San Antonio, Texas, USA.

Jaradat,D. and Tollefson ,M.(1988).The Impact of Alternative Scoring Procedures for Multiple-choice Items on Test Reliability, Validity, and Grading**. Educational and psychological Measurement**. 48, 627-635.

Keller, Lisa A.(2000). Ability Estimation Procedures IN Computerized Adaptive Testing. Retrieved May 13 2014 from: **www.aicpa.org** .

Kery, Marc.(2010). **Introduction to Winbugs For Ecologis**ts. Diacitec, Cennai, India.

Kong, Xiaojing.(2007). Using Response time and the Effort-Moderated Model to Investigate the effects of Rapid Guessing on Estimation of Item and Person Parameters. Unpublished Doctoral Dissertation, James Madison University, USA.

Haladyna, Thomas M.(2004). **Developing And Validating Multiple-Choice Test Items**. Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, USA.

Haladyna, Thomas M.; Downing, Steven M.(2004). Construct-Irrelevant Variance in High-Stakes Testing. **Educational Measurement: Issues and Practice**.

Hambelton, Ronald K. and Swaminthan, Hariharan. (1985). **Item Response Theory, Principles and Applications**. Boston, USA.

Hambleton, Ronald K and Cook, Linda L.(1983). Robustness of Latent trait models and Effect of test Length and Sample Size on the Precision of ability Estimates. **New Horizon in Testing**, 31-49.

Hambelton ,Ronald K.(1993). Item Response Theory: A Broad Psychometric Framework For Measurement Advances. **Psychometrika** , Vol 6,n 3.pp(353-556).

Hambleton,Ronald and Jones, Russel.(1994).Comparison of Classical Test Theory and Item Response Theory and their Applications to test Development. **Educational Measurement: Issue and Practice**.

Hambleton R.K, Cook.L.L.(1983). Robutness of item response models and effects of test length and sample size in the precision of ability estimate. NewYork.In D.J. Weiss(Ed),**New Horizons in Testing**,31-49.

Hambleton R.K. Swaminthan.H, and Rogers.H.J.(1991). Fundamentals of item response Theory**. International Educational and professional**. Sage Publication,Inc Newbury Park, USA.

Hambleton, Ronald. and Han, Kyung T.(2013). Wingen: A Proram for Generating Data [A Computer Software and Manual]. Retrieved 13/10/2013: from **http://www.mrc-bsu.cam.ac.uk**

Hambleton, Ronald. and Jones, Russell W.(1993). Item Paramter estimation errors and their influence on test information function. **Applied Measurment in Education**.7(3), 171-186.

Hambleton, Ronald K.(1989). Principles and selected Applications of Item Response Theory. In R. L. Linn(Ed), **Educational Measurement** 3th, 147-200. Macmillan New York, USA.

Han,Kyung T. 2012. Fixing the c Parameter in the Three-Parameter logistic Model. **Research and Evaluation**, 17,14-50.

Lesage, Ellen. Valcke, Martin and Sabbe, Elien.(2013). Scoring methods for multiple choice assessment in higher education- Is it still a matter of number right scoring or negative marking?.Studies in Educational Evaluation. 39(2013), 188-193.

Liao,Wen-Wei.(2012).The Four-Parameter Logistic Item Response Theory model AS Robust Method of Estimating Ability Despite Aberrant Responses. **Social Behavior and Personality** ,40,1679-1694.

Linden WimJ,and Hambleton,Ronald K.(1997). **Handbook of Modern Item Response Theory.** Springer-Verlage New York Inc,USA.

Lord, F. I. and Novick. M. R. (1968). **Statistical Theories of Mental Test Scores**. MA: Addison-Wesley.

Lord, F. M.(1975). Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters .(Research Bulletin 75-33). Princeton NJ: **Educational Testing Service**.

Lord,F.M.(1980).**Application of Item Response Theory to Practical Testing Problems**. Lawrence Erlbaum Associates, Hillsadle, New Jersey,USA.

Lord, Frederic M. (1984). Maximum Likelihood and Bayesian Parameter Estimation in Item Response Theory. Retrieved Dec 10 **http://files.eric.ed.gov**

Maraia,Paz. and Javier Gardeazabal.(2006).Optimal Correction for Guessing in Multiple-Choice Tests Assumptions. **Unpublished Doctoral Dissertation**, Brigham young Uinversity,USA.

Magis, David.(2013).A note on the item information function of the four –Paramter logistic model. **Applied psychological Measurement**, 37(4),304-315.

Mcmillan, James H.(2013). **Research on Classroom Assessment**. Sage Publication, USA.

Miller, M.David. Linn, Robert L. and Gronlund, Normane E.(2009). **Measurement and Assessment in Teaching**. 10th , Pearson Education, Inc,New Jersey,USA.

Messick,S.(1995).Validation of Psychological Assessment: Validation of Inferences from person's response and performances as scientific inquiry into scoring meanings. **American Psychologist**,50,741-749.

Mehrens,William A. and Lehmann,Irvin J.(1991). **Measurement and Evaluation in Education and Psychology**. 4th. Thomson Learning, USA.

Merrit, Stephen.(2006). **Mastering Multiple Choice: The Definitive Guide to Better Grades on Multiple Choice Exams**. Brain Ranch, Canada.

Muijtjens, A M M. Mameeren, H Van, Hoogenboom, R J I. and Evers J L.(1999). The Effect of a "do n't know " Option on Test Scores: Number- Right and Fiormula Scoring Copmared. **Medical Education**, 33, 267-275.

Natesan Prathiba. and Kieftenbeld,Vincent.(2012).Recovery of Graded Response Model Parameters:A Comparison of Marginal Maximum Likelihood and Markov Chain Monte Carlo Estimation. **Applied Psychological Measurement**. 36(5),399-419.

Nunnally, J.C.(1978).**Psychometric Theory**.New York,USA: Mccrow-Hill.

Ostini, Remo. and Nering, Michael L.(2006). **Polytomous Item Response Theory Models**. Sage Publication, Inc ,California,USA.

Partchev, Ivailo.(2004). **A Visual Guide to Item Response Theory. Friedrich-Schiller- University**. Jena, Germany.

Parshall, Cynthia G. Kromery, Jeffry D and Chason,Walter M.(1997). Evaluation of Parameter Estimation under Modified IRT Models and Small Samples. **Paper presented at the annual meeting of the Psychometric society**, Gatlinburg, TN, June 26-29.

Pelton, T.W.(2002). **The accuracy of Unidimensional Measurement Models in the Presence of Deviations for the Underlying Assumptions**. Unpublished Doctoral Dissertation. Brigham Young University, USA.

Poham, W.James.(2003). **Educational Assessment, Pearson Education**. Inc, USA.

Razali, Nornadiah. Razali, Wah. and Yap, Bee.(2011). Power Comparison of Shapiro-Wilk, Kolomogorov-Smirnov, Lilliefors and Anderson-Darling Tests. **Journal of Statistical Modeling and Analytics**, 2(1), 21-33.

Reckase,Mark D.(2009).**Multidimensional Item response Theory**. Springer Dordreclt Heidelberg London.

Reif,Manuel. Holocher-Ertl, Stefana. Klaus D, Kubinger,and Frebort,Martina.(2010). On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of two solutions and three Distractors item format to a one Solution and Five Distractors Item Format. **International Journal of Selection and Assessment**. 18(1), 110-115.

Rogers, H.J.(1999). Guessing in Multiple Choice Tests. In G. N.Master & J.P.Keeves(Eds),**Advances in Measurement in Educational Research and Assessment**, 15,235-243.

Rowley, Glenn L. and Traub, Ross E.(1977). Formula Scoring, Number- Right Scoring, and Test-Taking Strategy. **Journal of Educational Measurement**, 14(1), 15-22.

Sabbe Elien.Valcke, Martin. and Lesage Ellen.(2013).Scoring methods for multiple choice assessment in higher education is it still matter of number right scoring or negative marking?. **Studies in Education Evaluation**. 39,188-193.

San Martin, Ernesto. Del Pino, Guido.and De Boeck, Paul.(2006). IRT Models for Ability-Based Guessing.Applied Psychological Measurment, 30(3), 183-203.

Sireci, Stephen.(1992). The Utility of IRT in Small-Sample Testing Applications. Reterived April. 12.2013 from **http://eric.ed.gov**.

Shneider, M. Christian. Eagn, Karla L. and Julian, Marc W.(2013).Classroom Assessment in The Context of High-Stakes Testing. In McMillan. James H.(Ed), **Research on Classroom Assessment**. Sage Publication.

Setiadi,H.(1997).**Small Sample IRT Parameter Estimates**. Unpublished Doctoral Dissertation, University of Massachusetts, USA.

Setzer,J,Carl.(2007).**Studying the effect of Rapid-Guessing on low-Stakes test: An Application of the Effort-Moderated IRT Model**. Unpublished Doctoral Dissertation, James Madison University, USA.

Spiegelhalter, David. Thomas, Andrew. best, Nicky and Lunn, Dave.(2003). Windows Bayesian Using Gibbs Sampling [A Computer Program] Retrieved april.10.2013 from **http://www.mrc-bsu.cam.ac.uk.**

Sternberg, Robert J and Grigorenko, Elena L.(2002). **Dynamic Testing The Nature and Measurment of Learning Potential**. Cambridge University Press,United Kingdom.

Stocking, Martha L.(1990). Specifying Optimum Examinees for Item Parameter Estimation in Item Response Theory. **Psychometrika**, 55(3), 461-475.

Subedi, Dipendar Raj.(2009). **Investigating Unobserved Heterogeneity Using Item Response Theory Mixture Models**. Unpublished Doctorial Dissertation, Michigan University, USA.

Swaminathan, Hariharan and Gifford, Janice.(1983). Estimation of paramters in the 3-Parameter Latent Trait model. **New horizons in testing**, 373-385.

Chiu, Ting-Wei and Camilli, Gregory.(2012). Comment on 3PL IRT Adjustment for Guessing. **Applied Psychological Measurement** , 37(1), 76-86.

Thompson, Nathan A.(2009). **Ability Estimation with Item Response Theory**. Assessment Systems Corporation, USA.

Thissen ,David, and Wainer, Howard.(1983). Item Response Theory: Some standard Errors. **EST,Princeton**,NewJeresy, USA.

Traub,Rose E.(1977).Formula scoring, Number-right scoring, and test taking strategies. **Journal of Education Measurement**, 14,15-22.

Wade, L. Harrison, C. Hollands, J. Mattick, K, Ricketts, c. and Wass,L.(2011). Student perception of the prgores test in two setting and the implications for test deployment. Adv Health Sci Educ, Springer.

Wainer, L. and Wright,B.D.(1980). Robust estimation of ability in the Rasch model. **Psychometricka**, 45,373-391.

Wainer,Howard. Bradlow,Eric T. and Wang,Xiaohui.(2007).**Testlet Response Theory and its Applications**. Cambridge University Press,London.

Waller, M.I.(1989).Modeling Guessing Behavior: A comparison of two IRT Models. **Applied Psychological Measurement**, 13,233-243.

Waller, Niles. and Reise, Steeven.(2010).Measuring psychopathology With No standard Item Response Theory Models: Fitting Four-Parameter Model to the Minnesota Multiphasic Personality Inventory. **American Psychological Association**, Washington, DC,(147-173).

Wang, Shudong. And Wang, Tianyou.(2002). Relative Precision of ability Estimation in Polytomous CAT: A Comparison under the Generalized Partial Credit Model and Graded Response Model. **Paper Presented at the annual meeting of the American Educational Research Association**, New Orleans.

Warm, Thomas, A.(1978). **A Premier of Item Response Theory. Oklahoma**, USA: Us Coast Guard Institute.

Wise, S.L. and DeMars,C.E.(2005). Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions. **Educational Assessment**, 10,1-17.

Wise,S.L. and DeMars, C.E.(2006).An Application of Item Response Time: The Effort Moderated IRT model. **Journal of Educational Measurement**, 43,19-38.

Wise, S.L., and Kong,X.(2005).Response time effort: A new measure of examinee motivation in computer-based tests. **Applied Measurement in Education** ,18,163-183.

Weiss, David J. and Babcock, Ben.(2009). Termination Critreria in Comuterized Adaptive Tests: Variable-Length CATs Are Not Biased. Presneted at the Realities of CAT Paper Aession, Retrieved April 8.2015 from **http:// www.psych.umn.edu/psylabs/CATCentral.**

Xie,Yiyu.(2002).**An Application of a Special Two-Class Item Response Model Using Markov chain Monte Carlo Method**. Unpublished Doctoral Dissertation, University of California at Berkely,USA.

Yang,Xianggdong.(2007).Methods of Identifying Individual Guessers from Item Response Data. **Educational and Psychological Measurement**. 67(5), 20-34.

Yamamoto, K.(1995).Estimating the effects of test length and test time on Parameter Estimation using the Hybrid Model. Newjersy, USA:ETS.

Yamamoto, Ketaro.(1987). **A Model that Combines IRT and Latent Class Models**. Unpublished Doctoral dissertation, University of Illinois, USA.

Yen, Wendy. and Fitzpatrick, Anne R.(2006).Education Measurement .Inc. Brennan, Robert L, 4th **Educational Measurement**. .Praeger Publishers. USA.

Yu,Chong Ho.(2010). A simple guide to the item response Theory and Rasch Modeling. Retrieved March 13 2013 from :**http://www.creative-wisdom.com**.

ملحق رقم(١)

تقدير معالم الفقرات ومعلمة القدرة باستخدام برنامج (WinBugsv1.4)

تم استدعاء البيانات التي تم توليدها باستخدام برنامج (Wingen v.3) عند كل متغير من متغيرات الدراسة، وتمت عملية تقدير معالم الفقرات وقدرات المفحوصين بطريقة (Bayes) وفق الخطوات التالية: (الشواورة،٢٠١٣) ، (Kery, 2010).

١. إنشاء ملف أوامر نصي (Syntax) يلبي المتطلبات اللازمة لتشغيل كل نموذج من النماذج موضع الاهتمام.
٢. ملفات نصية تحوي استجابات المفحوصين وفق ظروف الدراسة من طول الاختبار واحجام العينات، حيث أن برنامج التحليل يشترط ان تكون البيانات في تنسيق محدد لكي يتم قراءتها من قبل البرنامج.
٣. النقر على قائمة Model من شريط اللوائح .
٤. النقر على الأمر"خصائص" (specification)
٥. النقر على زر "فحص النموذج" (Check model)
٦. تظهر رسالة نصية في سطر الحالة للبرنامج تفيد أن الاوامر المكتوبة صحيحة من ناحية نحوية (Model is Syntactically correct).
٧. الانتقال إلى نافذة البيانات من خلال قائمة النوافذ (Window)
٨. تظليل الكلمة المفتاحية (List) في ملف البيانات.
٩. الضغط على زر "تحميل البيانات" (Load data)
١٠. تظهر رسالة نصية في سطر الحالة تفيد أن البيانات قد حملت (Data Loaded)
١١. يتم النقر على زر "التوليف" (Compile)
١٢. تظهر رسالة في سطر الحالة مرة اخرى تفيد أن النموذج قد تم توليفه (Model Compiled).
١٣. يتم الضغط على زر " توليد القيم الابتدائية للتوزيعًات المسبقة" (Gen inits for prior Dis)
١٤. تظهر رسالة نصية في سطر الحالة تفيد انه "تم توليد القيم الابتدائية للتوزيعًات المسبقة"، وانه تم تجهيز النموذج (Initial Values generated model Initialized)
١٥. فتح قائمة Inference
١٦. اختيار الأمرSamples
١٧. ادخال الحروف الخاصة بالمعالم حسبما ذُكرت في ملف الأوامر مع مراعاة طبيعة النموذج، بمعنى يتم ادخال الحرف a في اشارة لمعلمة التمييز متبوعة بإجراء نقرة على زر التهيئة (set) فتدخل الحرف (b) في إشارة لمعلمة الصعوبة متبوعة بإجراء نقرة على زر التهيئة (Set) ثم ندخل رمز النجمة (*) في الحقل النصي للعقدة (node) متبوعة بإجراء نقرة على زر التهيئة (Set).
١٨. فتح قائمة النموذج Model مرة أخرى.

١٩. اختيار امر "التحديث" (Update)

٢٠. يظهر صندوق حواري (Interactive Dialog Box) عنوانه "أداة التحديث" (Update Tool)

٢١. وضع قيمة لعدد مرات التحديث في المرحلة التكيفية (Adaptive Stage).

٢٢. النقر على زر "التحديث" (Update)

٢٣. ستظهر رسالة نصية في سطر الحالة تفيد ان المرحلة التكيفية قد نفذت في عدد من الثواني.

٢٤. تحديد قيمة لعدد مرات التحديث (ويفضل ان تزيد عن القيمة المذكورة في النقطة(٢٢) لمرحلة الحرق(الاتلاف) (Burn In Stage).

٢٥. يتم النقر على زر التحديث.

٢٦. تظهر رسالة في سطر الحالة تفيد أن مرحلة التسخين(الحرق) قد نفذت في كذا ثانية.

٢٧. يتم وضع قيمة لعدد مرات التحديث (يفضل أن تكون من مضاعفات القيمة الموضوعة في النقطة (٢٥) لمرحلة تقدير المعالم بالشكل النهائي (Final Estimating Stage).

٢٨. تظهر رسالة نصية في سطر الحالة تفيد أن مرحلة تقدير المعالم النهائية قد نفذت في الزمن التالي، وهي مرحلة تستغرق الكثير من الزمن حيث وصلت إلى حوالي (٢٦) ساعة للنموذج الثلاثي وبحجم عينة(١٥٠٠) مفحوص.

٢٩. النقر على زر الاحصائيات (Stats) بهدف عرض نتائج عملية التقدير لمعالم الفقرات.

٣٠. من قائمة ملف (File) يمكن تخزين الاحصائيات في ملف خاص للتعامل معها في اتمام بقية اجراءات الدراسة.

ملحق رقم (٢)

الأوامر النصية للنموذج اللوجستي الثلاثي المعلمة

```
Model
{

# Read in individual item responses

for (i in 1:1500) {for (j in 1:50) {x[i,j] <- response[i,j]}}

#Identify two-parameter logistic (3PL) model

for (i in 1:1500) {for (j in 1:50) {

        e[i,j] <- exp(a[j]*(theta[i]-b[j]))
        p[i,j] <-  c[j]+(((1 - c[j])*e[i,j])/(1+e[i,j]))
        x[i,j] ~ dbern(p[i,j])
        mx[i,j] <- 1-x[i,j]
                                                    }
#Specify prior for examinee parameters

        theta[i] ~ dnorm(0,1)
            }
#Specify priors for item parameters

for (j in 1:50) {a[j] ~ dlnorm(0,2) b[j] ~ dnorm(0,1) c[j] ~ dbeta(1,5)}


# End of model
}
list(response=structure(.Data=c(
0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1,
1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1,
0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1,
0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1,
0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1………,
```

ملحق رقم (٣)

الأوامر النصية لنموذج عتبة التخمين

```
for (i in 1:N}(

        for (j in 1:J}(

logit(p[i,j]) <- gamma[j]*(theta[i]-delta[j]) -step(j-alpha[i]-.5)*
   (gamma[j]*(theta[i]-delta[j]) -k[j])

         x[i,j] ~ dbern(p[i,j])

    {

        alpha[i] ~ dcat(PI([])

   {

    For (j in 1:J)

      delta[j] ~ dnorm(0,1)

      gamma[j] ~ dgamma(1,1)

      k[j] ~ dbeta(8, 32)

   {

    for(i in 1:N){

      theta[i] ~ dnorm(0,1)

   {

    PI[J] ~ dbeta(1,1)

    temp1 <- ( 1.0 - PI[J] ) / pow((J-1),

    PI[1] <- temp1

    for(j in 2:(J-1}((

      PI[j] <- (pow(j, b)-pow((j-1), b))*temp1

   {

    b ~ dgamma(2,1)
```

ملحق رقم (٤)

الأوامر النصية للنموذج الثنائي المُعدّل

```
Model
{

# Read in individual item responses

for (i in 1:500) {for (j in 1:50) {x[i,j] <- response[i,j]}}

#Identify two-parameter logistic (3PL) model

for (i in 1:500) {for (j in 1:50) {

e[i,j] <- exp(a[j]*(theta[i]-b[j]))
p[i,j] <-  .20+(((1 - .20)*e[i,j])/(1+e[i,j]))
x[i,j] ~ dbern(p[i,j])
mx[i,j] <- 1-x[i,j]
                                                    }
#Specify prior for examinee parameters
theta[i] ~ dnorm(0,1)
      }
#Specify priors for item parameters

for (j in 1:50) {a[j] ~ dlnorm(0,2) b[j] ~ dnorm(0,1) c[j] ~ dbeta(1,5)}


# End of model
}
```

ملحق رقم (٥)

الأوامر النصية للنموذج القائم على الصعوبة

```
Model
{ for (i in 1:1500){
     for (j in 1:50){

         logit(p[i,j]) <- gamma[j]*(theta[i]-delta[j]) -(alpha[i]-
1)*step(delta[j]-theta[i]-b)*(gamma[j]*(theta[i]-delta[j]) -k[j])
         x[i,j] ~ dbern(p[i,j])
          }
     alpha[i] ~ dcat(PI[])
     }

  for(j in 1:50){

     delta[j] ~ dnorm(0,1)
     gamma[j] ~ dgamma(1,1)
    k[j] ~ dbeta(3, 8)
     }
  for(i in 1:1500){
     theta[i] ~ dnorm(0,1)
     }
  PI[1] ~ dbeta(1,1)
  PI[2] <- 1.0 - PI[1]
  b ~ dnorm(0,2)
  taub ~ dgamma(0.001, 0.001)
 }
```

ملحق رقم (٦)
شاشة برنامج (Winbug)



Licence Agreement

Introduction

This software and any associated documentation whether electronic or printed (hereinafter called "WinBUGS PACKAGE") is made available under a licence agreement and may be used only in accordance with the terms of that agreement.
This is a legal agreement between you (the Licensee), and MRC and Imperial College of Science, Technology and Medicine (the Licenser). The terms of the licence are provided in the following pages.

Users are required to register and to pay a fee for the use of the WinBUGS PACKAGE. Details of fees and the procedure for registration and acceptance of the licence terms is provided here.
There is no fee payable for the use of the demonstration (Internet) version of the WinBUGS package. Users of the demonstration version of the WinBUGS package can upgrade to the full version on payment of a fee.

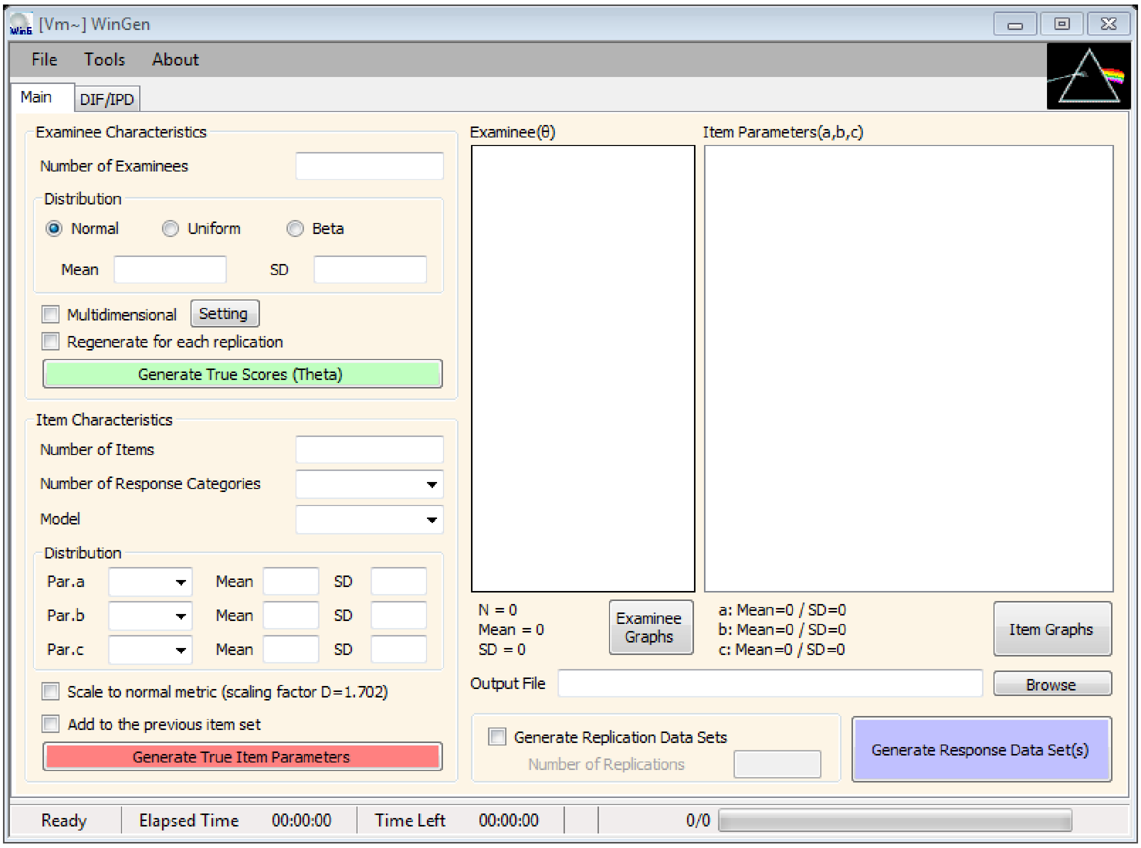
The current fee is zero dollars ($0).

By completing and sending the registration you demonstrate your agreement to the terms of this licence and will become legally bound to the terms therein.

It should be emphasised that the statistical tools provided in the WinBUGS PACKAGE are by their very nature partly subjective. The Licenser cannot offer advice on interpretation of results obtained using the WinBUGS PACKAGE.
Any assistance will be strictly limited to attempting to help if there are problems formulating the statistical problem with the WinBUGS PACKAGE.

The Licencer makes no representation or warranties with respect to the WinBUGS

ملحق رقم (٧)
شاشة برنامج (Wingen3)



COMPARING THE ACCURACY OF FOUR MODELS OF HANDLING GUESSING IN MULTIPLE CHOICE ITEMS WITHIN THREE PARAMTERS LOGESTICS MODEL ON ESTIMATING INDIVIDUALAS ABILITY AND ITEM PARAMTERS

By

Moueen Slman S. Al- Nsraween

Supervisor

Dr. Mohamed W. Al-Batsh, Prof

ABSTRACT

This study investigated the efficiency of four item response models which are used to estimate examinee ability while correcting for guessing in multiple choice items. The data used varied in test length (25 and 50 items), and examinee sample size (500 and 1500 examinees). The four models investigated were: the three parameter logistic model, the two parameter modified model with a rate of guessing fixed at (0.20) for all persons, the threshold guessing model, and the difficulty-based guessing model.

A simulated examinee-response dataset was generated by fitting a three parameter logistic model using (Wingen3) program, with the specified ability and difficulty parameters distributed normally with mean of (0) logits and standard deviation of (1) logits. The discrimination parameters were distributed uniformly with their initial value set equal to (0.40) and a maximum-possible value of (1.7) The guessing parameters were generated uniformly with initial value of (0) and a maximum value of (0.45).

The unidimensionality of the generated data was investigated using two procedures: internal consistency magnitudes using Cronbach alpha, and the factor analysis of the item covariance matrix using principle components analysis. In order to evaluate the shape of the distribution of the generated ability and difficulty parameters, the Kolomogrov-Smirnov test was

computed and evaluated for data-concordance with normal distribution expectations. Bayesian estimation (using the Winbug software) was used for estimating item parameter and ability parameters within the four models used in this study to estimate/correct-for guessing.

In order to evaluate the estimation efficiency and accuracy of the four models deployed in this study, the estimated item and ability parameters were analyzed using the Root Mean Square Standard Error (RMSE), Relative-Efficiency, Bias, and examining the mean and standard deviation for the standard error of estimate. A repeated measurement design was conducted for these investigations, using SPSS statistical Package for social science.

The study results showed statistically differences between (RMSE) means for ability parameter estimation at ($\alpha = 0.05$) when using the three logistic parameter model compared with the other three models. This indicates the superior accuracy of the three-parameter logistic model. In addition, the results demonstrated statistically-significant differences at ($\alpha = 0.05$) between means for (RMSE) , and (Bias), when estimating item parameters (discrimination, difficulty, and guessing). The relative-efficiency indicator also demonstrated the same superiority for the three logistic parameter model and the two-parameter modified model, with the results showing that the means for standard error of estimate, RMSE, and Bias, decreased as the sample size increased from 500 to 1500 examinees and as test length increased from 25 to 50 items. **Key words: Guessing, Accurate Estimation, Modified Models, Mixture Models.**